‘Measuring well-being for public policy’

Tessa May Peasgood

Submitted for Doctorate in Philosophy

Tanaka Business School

Imperial College London

2008
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Paul Dolan, without whose encouragement I would not have begun this thesis. I have both enjoyed and benefit immensely from our many happy hours of discussion and debate.

Thanks also to Ian White for, amongst many other things, insightful comments on parts of this thesis, strengthening my arguments and reducing my occasionally verbose text.

Thanks to Jennifer Roberts for help and discussion in relation to the econometrics. Robert Metcalfe, Jenny Brazier and Anna Peasgood for helpful comments and proof reading.

Finally, thanks are due to other researchers in the field with whom I have enjoyed interesting and inspiring discussions.

The data used in this thesis were made available through the ESRC Data Archive. The data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex (now incorporated within the Institute for Social and Economic Research). Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here.
ABSTRACT

The aim of this thesis is to explore the most appropriate means of measuring individual well-being for public policy. The thesis will begin by asking exactly what is meant by well-being and giving a brief overview of the main philosophical accounts of well-being, summarising the variations within each account and considering how well-being can be measured from that perspective. Chapter Three uses analysis of the British Household Panel Survey (BHPS) to explore the differences between the well-being measures available within that dataset. This shows that, although in many cases the choice between different subjective well-being (SWB) measures appears to make minimal difference, there is wide variation between household income and subjective well-being measures and also some important cases where subjective measures differ. This implies that the choice of well-being measure will have important implications for policy making.

To aid judgement between different well-being measures a list of criteria for an ideal measure of well-being to be used for public policy are set out in Chapter Four. Chapter Five asks how the various measures of well-being fair according to these criteria, finding that subjective measures perform at least as well, if not better, than existing measures of well-being drawn from neo-classical economic theory.

The next three chapters use evidence from the SWB literature and additional analysis of the BHPS to asks whether our current understanding of the determinants of SWB is sufficient to support policy recommendations. Chapter six focuses on the relationship between income and SWB, specifically on evidence for inefficiencies arising from adaptation to income and reference income effects. Chapter seven looks at the relationship between SWB and social capital and considers the extent to which the relationship between talking to neighbours and SWB can be thought of as causal through the use of Instrumental Variable techniques. Lastly, chapter eight considers the role of SWB measures within microeconomic evaluation. The final chapter summarises the advantages and potential hurdles for the use of SWB indicators of well-being within public policy, concluding with recommendations for extending this work, both in relation to public policy and academic research.
CONTENTS

Chapter 1: Introduction ............................................................................................................. 1

1.1 Introduction ......................................................................................................................... 1

1.2 Government policy making needs a measure of individual well-being .................................. 2

1.3 Current measures of well-being are not fit for policy .......................................................... 2

1.3.1 The inadequacy of welfare economics .............................................................................. 2

1.3.2 Current use of welfare economics measures .................................................................... 10

1.3.3 Income as a measure of well-being appears to lack convergent validity ......................... 12

1.3.4 Problems with other measures .......................................................................................... 13

1.3.5 The current state of measures of well-being .................................................................... 15

1.4 Reconsider measures of well-being ..................................................................................... 15

1.4.1 The UK Government is calling for a new measure of well-being .................................. 15

1.4.2 New measures of well-being from psychology ................................................................. 16

1.4.3 Summary of the need to reconsider measures of well-being ........................................... 17

1.5 What type of measure of well-being is required? .................................................................. 17

1.6 Outline of the thesis, methodology and contribution ......................................................... 18

Chapter Two: Measuring well-being .......................................................................................... 24

2.1 Introduction ........................................................................................................................ 24

2.2 Preference accounts ............................................................................................................ 25

2.2.1 Theories of preference or desire satisfaction .................................................................. 25

2.2.2 Measuring desire satisfaction ......................................................................................... 26

2.3 Mental state accounts ......................................................................................................... 35

2.3.1 Hedonic accounts ........................................................................................................... 36

2.3.1.1 Theories of hedonism ................................................................................................. 36

2.3.1.2 Measuring hedonic well-being ................................................................................ 37

2.3.2 Evaluative accounts ........................................................................................................ 43

2.3.2.1 Evaluative theories of well-being ............................................................................. 43

2.3.2.2 Measuring evaluative accounts ................................................................................ 47

2.3.3 Subjective well-being ..................................................................................................... 51

2.3.3.1 Theories of subjective well-being .............................................................................. 51

2.3.3.2 Measuring subjective well-being .............................................................................. 53

2.4 Hybrid accounts ............................................................................................................... 57

2.5 Summary and discussion ................................................................................................... 58

Chapter Three: Empirical differences between measures of well-being ..................................... 63

3.1 Introduction ....................................................................................................................... 63

3.2 Evidence from the literature .............................................................................................. 63

3.2.1 Income as a proxy of utility versus SWB ........................................................................ 63

3.2.2 Various subjective measures of well-being ..................................................................... 64

3.3 Well-being measures available in the BHPS ...................................................................... 67

3.3.1 Income and consumption .............................................................................................. 67

3.3.2 Health-related quality of life ......................................................................................... 71

3.3.3 CASP-19 ........................................................................................................................ 71

3.3.4 SWB past and future ..................................................................................................... 72

3.3.5 Life satisfaction measures ............................................................................................. 72
3.3.6 Depression and anxiety scale GHQ ................................................................. 79
3.4 Variability between different well-being measures in the BHPS ....................... 80
3.5 Are differences in well-being measure related to individual characteristics? ........ 82
3.6 Differences in the factors associated with well-being ..................................... 90
3.7 Summary ............................................................................................................. 102
APPENDIX 3.A ............................................................................................................ 104
APPENDIX 3.B ............................................................................................................ 106
General Health Questionnaire 12 (GHQ12) Scale .................................................. 106
CASP-19 Scale ........................................................................................................ 107
APPENDIX 3.C ............................................................................................................ 108
APPENDIX 3.D ............................................................................................................ 110
APPENDIX 3.E ............................................................................................................ 118
APPENDIX 3.F ............................................................................................................ 127

Chapter 4: Criteria to judge whether a measure of well-being can be useful for public policy .. 130
4.1 Introduction ......................................................................................................... 130
4.2 Choosing the assessment criteria ...................................................................... 131
4.2 Appropriate conception of well-being for public policy ..................................... 137
4.2.1 Prudential value only ...................................................................................... 137
4.2.2 Inter-temporally and interpersonally comparable .......................................... 138
4.3 Validity ............................................................................................................... 139
4.3.1 Content validity .............................................................................................. 139
4.3.2 Convergent validity ......................................................................................... 140
4.3.3 Predictive validity ............................................................................................ 140
4.4 Empirically useful .............................................................................................. 140
4.4.1 Reliable with low and unbiased measurement error ....................................... 140
4.4.2 Cardinality ...................................................................................................... 141
4.4.3 Appropriate sensitivity ..................................................................................... 142
4.4.4 Practical and efficient to collect ...................................................................... 143
4.4.5 Public/political acceptability ......................................................................... 143
4.5 Conclusion ......................................................................................................... 145

Chapter 5: How closely do well-being measures meet the ideal criteria? ..................... 147
5.1 Introduction ......................................................................................................... 147
5.2 Appropriate conception of well-being for public policy ..................................... 147
5.2.1 Prudential value only ...................................................................................... 147
5.2.2 Inter-temporally and interpersonally comparable .......................................... 149
5.3 Validity ............................................................................................................... 158
5.3.1 Content validity .............................................................................................. 158
5.3.2 Convergent validity ......................................................................................... 164
5.3.3 Predictive validity ............................................................................................ 166
5.4 Empirically useful .............................................................................................. 168
5.4.1 Reliable with low and unbiased measurement error ....................................... 168
5.4.2 Cardinality ...................................................................................................... 172
5.4.3 Appropriate sensitivity ..................................................................................... 177
5.4.4 Practical and efficient to collect ...................................................................... 179
5.4.5 Public/Political acceptability ......................................................................... 181
5.5 Conclusion ......................................................................................................... 182
TABLES

Table 1: Domain satisfactions, wave 14 .................................................. 72
Table 2: Coefficients on domain satisfactions from GLS-RE & OLS-FE on life satisfaction .............. 77
Table 3: Coefficients on domain satisfactions from an OLS-FE on life satisfaction, by subgroup ........ 78
Table 4: Coefficients for education level from FE models on life satisfaction .................................. 99
Table 5: Coefficients for education level from RE models on life satisfaction ..................................... 100
Table 6: Education coefficients from a generalised ordered probit RE model on life satisfaction ..... 101
Table 7: Correlations between income and consumption, wave 14 .............................................. 104
Table 8: Correlations between well-being measures, wave 14 ..................................................... 105
Table 9: Correlations between well-being measures, wave 11 ..................................................... 105
Table 10: Description of variables, waves 7-14, unless stated ......................................................... 110
Table 11: Measures of well-being, waves 7-14, unless stated ....................................................... 116
Table 12: OLS coefficients on various well-being measures, wave 11 ........................................... 118
Table 13: OLS coefficients on various well-being measures, waves 9 & 14 ................................. 121
Table 14: OLS-FE coefficients on various well-being measures, waves 9 & 14 .............................. 124
Table 15: Studies finding a positive impact of education on SWB .................................... 127
Table 16: Studies finding a negative impact of education on SWB .................................... 128
Table 17: Studies finding a non-significant impact of education on SWB............................. 128
Table 18: Studies showing the impact of education on mental health.................................. 129
Table 19: Possible interpretation of a 1-7 life satisfaction scale ........................................ 175
Table 20: Estimates of temporal stability of SWB measures .............................................. 184
Table 21: OLS-FE on life satisfaction, including past income ........................................... 193
Table 22: OLS-FE on life satisfaction, including past income, for those with and without a mortgage
                                                                                         ....................................... 194
Table 23: OLS-FE on income satisfaction and perceptions of financial coping ....................... 194
Table 24: OLS-FE on life satisfaction, for those with rising and falling incomes.................... 195
Table 25: OLS-FE on life satisfaction, including future expectations .................................. 195
Table 26: GLS-RE and OLS-FE on life satisfaction, including the importance of income ........ 196
Table 27: GLS-RE on the relative importance of income .................................................. 197
Table 28: OLS-FE on life satisfaction, including relative income ....................................... 213
Table 29: OLS-FE on life satisfaction, including past and relative income............................ 213
Table 30: OLS-FE on life satisfaction, including relative income, for ‘movers’ and ‘non-movers’ ... 214
Table 31: OLS-FE on life satisfaction, including income inequality .................................... 215
Table 32: OLS-FE on life satisfaction, by age group .......................................................... 215
Table 33: OLS-FE on life satisfaction, income relative income and income inequality, but not controlling for income ................................................................. 216
Table 34: Tobit on monthly savings, for the head of the household ..................................... 218
Table 35: FE Conditional logit and RE logit on ‘debts being a problem’ ................................ 220
Table 36: Hypothetical example showing incentives to change work hours ............................ 225
Table 37: Prisoner’s dilemma of work hours under the negative externality of relative income... 226
Table 38: OLS-FE and IV-FE for life satisfaction, including talking to neighbours.................. 245
Table 39: OLS-FE and GLS-RE on ‘talking to neighbours’ ................................................ 246
Table 40: Regressions on life satisfaction and domain-weighted satisfaction ......................... 263
Table 41: Regressions on life satisfaction and domain-weighted satisfaction, including sf-6d .... 264

FIGURES
Figure 1: Mean standardised score by well-being measure for the over 70s.............................. 83
Figure 2: Mean standardised score by well-being measure for men ....................................... 84
Figure 3: Mean standardised score by well-being measure for those with a degree ................ 85
Figure 4: Mean standardised score by well-being measure for the unemployed ...................... 86
Figure 5: Mean standardised score by well-being measure for the disabled ........................... 87
Figure 6: Mean standardised score by well-being measure for single parents ......................... 88
Figure 7: Mean standardised score by well-being measure for those with children ................. 88
Figure 8: Mean standardised score by well-being measure for those from London ................. 89
Figure 9: Mean standardised score by well-being measure for those who commute at least one hour/day ...............................................90
Figure 10: Coefficients for income on a range of SWB dependent variables.................................93
Figure 11: Coefficients for having a degree on a range of SWB dependent variables.........................94
Figure 12: Coefficients for being male on a range of SWB dependent variables..............................95
Figure 13: Histograms of well-being measures .................................................................................108
Figure 14: Frequency of satisfaction with life, BHPS wave 14, with mean (sd) of talking to neighbours scale (1-4) .........................................................................................................................233
Figure 15: Frequency of talking to neighbours, BHPS wave 14, with mean (sd) life satisfaction (1-7 scale) .................................................................................................................................233
Chapter 1: Introduction

1.1 Introduction

Well-being, broadly conceived, is how well an individual’s life is going. However, there are many different ways of conceptualising and measuring well-being. If all measures presented a similar picture of individual well-being then the choice of measure for policy would have minimal impact. Unfortunately for policy makers, this is not the case.

This thesis analyses data from the nationally representative British Household Panel Survey (BHPS) to show that measurement choice matters. The BHPS contains a range of different measures of well-being, including income, consumption, life satisfaction, health-related quality of life and mental health, all of which will be discussed in detail in Chapter Three.

Taking seven currently used measures of well-being, the BHPS data shows that in 2004 no one fell into the top quintile for every well-being measure and less than 2% fell into the bottom quintile for all measures. If the answer to ‘what is i’s well-being?’ varies dramatically depending upon the measurement instrument, this has ramifications for the use of well-being measures.

I have three aims in this thesis. First, to consider the theoretical and empirical differences between well-being measures and to understand the implications of these differences for policy. Second, to assess different well-being measures regarding whether they offer an ideal measure for policy purposes. This evaluation identifies subjective well-being (SWB) measures as useful for policy. Consequently, the third aim is to consider exactly how SWB can be used by policy makers. I consider the strength of evidence derived from analysis of SWB data and whether the evidence is sufficient to draw policy conclusions. Finally, I identify gaps in our knowledge of SWB measures requiring further research.

This chapter justifies the exploration of different well-being measures. I will demonstrate that in the interests of attaining an efficient and equitable allocation of resources a more accurate measure of individual well-being is necessary.

The value of this exercise depends upon two assumptions: governments require a measure of well-being to inform policy making; and current measures of well-being are inadequate. These two points are addressed below.
1.2 Government policy making needs a measure of individual well-being

There are three typical stances a government could take regarding the well-being of its citizens: a traditional welfarist approach, where welfare maximisation is the primary objective; a “weak welfarism” (Adler and Posner, 2006), where well-being is one of various considerations of action; and a libertarian, or rights, based approach where well-being is not an important consideration.

Strong and weak welfarism both require a measure of well-being, as their aim is to maximise well-being, either alone or in conjunction with other goods. Within the libertarian position, there is no obvious need to measure well-being, as the focus is on strengthening rights and freedoms. Robert Nozick (1974) for example, argues for the primacy of natural rights regardless of the consequences for human welfare. An individual’s right to have private property and bodily integrity protected can exist independently of knowledge of the impact this will have on her well-being.

Probably most libertarian-governments are actually weak welfarists in practice and most Western governments explicitly use measures of well-being in policy formulation. For example, the Scottish Executive Social Research report on quality of life notes that, “Virtually every realm of public policy making and service delivery in advanced capitalist nations is now influenced by notions of Quality of Life (QoL) and well-being” (Scottish Executive Social Research, 2005: 9). Hence, although a pure libertarian would not be interested in the debate over measuring well-being, few governments adopt that perspective.

Governments currently use measures for both individual well-being and social well-being or a measure of the success of society as a whole. In order to fully explore distributional issues and provide a measure of individual well-being to be used in economic evaluation, individual level measurement will be necessary. This work will focus only on well-being at the individual level.

1.3 Current measures of well-being are not fit for policy

1.3.1 The inadequacy of welfare economics

Background
Utility, or welfare, was traditionally conceived as a mental state like ‘happiness’ or ‘pleasure’ (Bentham, 1907). Jevons clearly saw the purpose of economic theory as being to support the goal of maximising happiness. His theory is “entirely based on a calculus of pleasure and pain; and the object of Economy is to maximise happiness by purchasing pleasure, as it were, at the
lowest cost of pain” (Jevons 1866: 27). However, he also perceived no means by which to directly measure happiness (Jevons, 1971: 14).

Concerns over the impossibility of directly measuring happiness led welfare economics to shift its roots away from maximising pleasure to the notion that an individual’s life goes better for her if she has more of her desires met. Under the assumption of perfect information and rationality (the individual will always choose that which maximises her utility defined as happiness) these two approaches converge. The assumption of rationality dodges many difficult normative questions raised by trying to measure preferences or agree a particular conception of preference satisfaction.

In his influential 1932 essay, Lionel Robbins casts economics as a tool to aid individuals in achieving their own ends, thereby removing concern over whether those ends were in a normative sense good for the individual. Human ends themselves were “taken for granted” and exploring the motivation behind those ends was taken to be both unnecessary and beyond the remit of economics (Robbins, 1932 [1945]: 24). Actual preferences held by individuals are therefore treated as exogenous, removing the need to worry about changes in preference and allowing focus only on changes in the number of desires met.

Paul Samuelson (1938) aimed to focus the discipline upon only that which is observable, namely choices made by individuals. If an individual chooses A over B, when both A and B are available, it can be said that the individual prefers A to B. Therefore, the individual gains more utility from A, where utility refers to being chosen. Hence from an individual’s actual choices (their revealed preferences), it is possible to know about their preference orderings between goods and states of the world.

This implies that an individual’s life goes better for her the more of her preferences are met. However, how preferences were to be aggregated into an overall level of well-being was not taken as an important aspect of this theoretical development, the focus being on marginal decisions (the additional utility from the choice of A over B) as oppose to the aggregate of all choices. Ordinal utility was taken as sufficient to analyse consumer behaviour, and the sacrifice of even a theoretical ability to aggregate complete individual well-being or aggregate individual well-being to societal well-being was perceived as necessary in order to achieve a more rigorous and mathematically defensible theory of consumer behaviour. Indeed Samuelson aimed to develop a theory of consumer behaviour “freed from the vestigial traces of the utility concept” (Samuelson, 1938: 71), and which was removed from what he saw as the ‘redundant’ concerns

1 Although it is possible to distinguish between preferences and desires, this work uses these words interchangeably.
of welfare economics. ‘New welfare economics’ aimed to be positive, and avoid troublesome normative judgements. The apparent advantages of this are neatly put by George Archibald:

“We need not encumber ourselves with meta-physical inquiries into the "true meaning" of "happiness" or "satisfaction" or "utility ". We need not preface our theorems with irrelevant value judgments. Instead, we proceed directly to our enquiry: under what arrangements will this individual’s choice be expanded? What are the necessary conditions for a situation in which no one’s choice may be expanded without a contraction in someone else’s?” (Archibald, 1959: 319-320).

If overall desire satisfaction increases more when a strongly held desire is realised than when a weakly held desire is realised, then the extent of any increase in well-being will depend on the strength of the desire for the goods and services attained. The relative strength of desire for a good, service or activity can be assumed to be reflected in the price individuals are willing to pay.\(^2\) If market price is a good indication of willingness to pay and willingness to pay (WTP) is a good indication of strength of preference, then additional expenditure on goods and services represents additional utility or well-being attained.

Assuming an individual’s preferences remain static, changes in well-being could therefore be inferred from changes in expenditure or consumption. Despite the fact that few economists would argue that in all circumstances where an individual’s preferences are satisfied their well-being is enhanced, or that consumption is an accurate measure of the extent of preferences fulfilled, consumption and income have still been taken to be the closest thing available to a measure of well-being.

Welfare economics assumes that an individual’s choices can be revealed through their market behaviour or stated choices, that choices mirror preferences, and that the satisfaction of preferences enhances well-being. This assumes individuals have consistent, non-altruistic, preferences which are complete \( ((y_1, y_2) \geq (x_1, x_2)) \text{ or } ((y_1, y_2) \leq (x_1, x_2)) \), reflexive \( ((x_1, x_2) \geq (x_1, x_2)) \), and transitive \( (x_1, x_2) \geq (y_1, y_2) \text{ and } (y_1, y_2) \geq (z_1, z_2) \text{ then } (x_1, x_2) \geq (z_1, z_2)) \).

**Challenges to welfare economics**

Welfare economics has attracted much criticism, particularly in recent years. There is a growing body of evidence that threatens to undermine the credibility of both using a preference satisfaction account of well-being (in which an individual’s life goes better for her if she gets

\(^2\) Samuelson (1974) used the term ‘direct money metric utility function’ for the minimum expenditure at a certain price level necessary to purchase bundles of goods at least as desired as an alternative bundle of goods.
what she wants), which is the concept of well-being underlying welfare economics, and using income as a measure of preference satisfaction.

Criticisms of welfare economics have centred on doubts that an individual’s choices are a good reflection of her preferences, and doubts that satisfying an individual’s preferences really maximise her well-being.

Evidence from behavioural economics and psychology has found that our preferences are not necessarily stable over time. This work has shown a range of factors that make it unlikely that people will always choose what is in their long-term interests. These include misinformed preferences (such as inaccurate predictions about the future, our inability to predict adaptation to changes in circumstances and the influence of our affective response to immediate situations) and preferences which do not lead to maximising behaviour over the long term (such as those due to complex interactions between current consumption of a good and the future benefit arising from future consumption of the good, a desire to rationalise our decisions which leads to a focus on partial outcomes, low will power and a tendency to heavily discount future benefits). These factors will be briefly discussed.

Our preferences may be misinformed because we inaccurately predict the future. There are various sources of information that we may use when predicting the future, many of which contain sources of error. Firstly we may look to our past experience, however, this information may be biased by our faulty recall (Kahneman, 1999; Rabin, 1998: 30), and a tendency to re-interpret the past (Mitchell et al., 1997). Secondly, we make assumptions about the future that do not adequately predict our future preferences in changed circumstances, particularly where is likely to be adaptation (Gilbert et al., 1998; Wilson et al., 2000; Wilson and Gilbert, 2003). One of the reasons why we fail to recognise the importance of a range of adaptation processes, expecting both good and bad feelings to last for longer, is that we fail to appreciate our ability to ‘make sense’ of things that happen to us (Wilson and Gilbert, 2003). People also fail to anticipate changes in their preferences caused by ownership (Kahneman et al., 1991), despite the “vast experience most people have had acquiring, possessing and losing objects” (Loewenstein and Adler, 1995).

At the time of making a decision, our actual choices are synonymous with our current preferences, where preference means to be in a chosen position. However, actual choices may diverge from preferences held prior to the moment of choice. Wilson et al. (2000) claim that our choices often reflect our immediate affective reactions to the choice we are faced with, rather than our underlying preferences. For example, when shopping for food people tend to shop as is
they expect to remain hungry (Gilbert et al., 2002). If preferences are not stable over time it will not be possible to derive a clear order of alternatives.

There is also evidence to suggest that we fail to predict how we will respond to temptation and that some choices are based on erroneous assumptions about our willpower. The large number of credit card users who incur high interest rates and charges has been viewed as implying that many credit card users expect to maintain a zero balance but fail to do so (Ausubel, 1991).

Even under well-informed preferences maximising over the long-term has been shown to be highly unlikely. People may suffer from low will-power. People may also make current choices according to which choice yields the highest utility in the moment without taking full account of the impact that choice will have on the utilities from future choices. This process, known as melioration, arises when people ignore ‘internalities’, or the impact that current consumption will have on the benefit of future consumption. These future impacts may be negative as in the case of an addictive substance, which can lead to “a devastating level of overindulgence in some commodity or activity” (Herrnstein and Prelec, 1991: 149), or they may be positive, as in the case of activities whose value increases with consumption such as skilled activities like playing a musical instrument. Where the utility of current consumption will depend upon past consumption, achieving optimum levels of consumption in the current time period involves both a good understanding of the melioration effects and accurate knowledge of future consumption.

In a similar vein, people’s choices may be based on excessive discounting or myopia. This is conceptually different from melioration since melioration could exist without time discounting because people are unable to assess correctly and compute the costs of actions (e.g. increase tolerance, degradation of other activities).

Our choices may diverge from those which maximise well-being because of a desire to rationalise our decisions and have clear reasons for making decisions, which leads people to focus on aspects which are easy to articulate (Hsee et al., 2003).

Furthermore, there are many areas in which individuals are unable to ‘reveal’ their preferences and do not have the option of choice. An individual can only choose what is under their control, “she can choose to work harder and consume more private goods but cannot on her own choose to have less pollution with less consumption for every person” (Ng, 1997: 3). Consequently, a more exhaustive measure of preference satisfaction requires incorporating hypothetical stated preferences, which results in additional theoretical and empirical problems (Carson et al., 2001; Loomes, 2006).
This evidence, much of which has accumulated in the last 15 years, presents a serious theoretical challenge to welfare economics. Challenges are unproblematic if they do not undermine the usefulness of the theory, often because they highlight exceptional or unusual cases. It has always been accepted that there will be cases in which choices do not adequately reveal preferences, and some preferences that would not maximise individual well-being. As noted by Little:

“.. we think that some people consistently fail to realize what would make them happy or happier, and to say of such people that they are in a chosen position, or on a higher behaviour line, may not imply that we think they are in any sense better off… I think most people would be prepared to admit that for a normal consumer our criterion was, in general, a good one. They would, in fact, accept that those people who are choosers, that is, excluding lunatics and children, know well enough how to order their own affairs to their best advantage. But we must recognize that again most people would think that there are some exceptions (Little, 1950: 45).

Such exceptions do not present a problem for the theory, as long as on average, most people, most of the time, act in such a way that these assumptions are reasonable.

Whilst criticisms may not be such that they undermine the credibility of theory, they appear to have reached a point where a re-assessment of the use of revealed and stated preference as a measure of value is needed. Behavioural economics has found behaviour, in both laboratory settings and the real world, which suggests violation of preference coherence, such as incompleteness, intransitivity and dynamic inconsistency. Instability of preferences and failure to maximising preference satisfaction due to lack of information and cognitive biases represents a serious challenge to the use of preference satisfaction as a measure of well-being and the use of WTP to represent changes in value.

Another major concern with the use of income as a proxy for well-being is the lack of consensus over the marginal utility of income. Although it is widely accepted that there is diminishing marginal utility of income, researchers are no closer to agreeing on the extent to which utility diminishes with income.

*The elasticity of the marginal utility of income is uncertain*

Economists have typically assumed that there is some diminishing marginal utility to income (Gossen, 1854). An additional £100 at £100,000 annual income meets fewer additional preferences than £100 at £10,000 annual income. Consequently, a poor person who is willing to pay £x for a good is likely to value this good more highly than a rich person willing to pay the same amount. In order to convert WTP amounts into the value attached to a good, we therefore need to know the marginal utility of income. However, finding out the nature of the relationship between additional income and additional utility at different levels of income is difficult, and no
consistent picture has arisen on how the marginal utility of income changes over levels of income. Consequently, there has been minimal effort to incorporate diminishing utility of income when considering changes in well-being at the individual level.

The marginal utility of consumption \( \frac{dU}{dC} \) is assumed to be positive (more consumption increases utility), but strictly decreasing in consumption due to diminishing marginal utility (more consumption increases utility at a decreasing rate). Hence,

\[
U'(C) = \frac{dU}{dC} > 0 \quad \text{and} \quad U''(C) = \frac{d^2U}{dC^2} < 0
\]

The elasticity of the marginal utility of consumption \( e \) measures the percentage rate at which marginal utility changes (falls) for every percentage change (increase) in consumption.

\[
e = \frac{CU''(C)}{U'(C)} > 0
\]

The elasticity of the marginal utility of consumption \( e \) may depend upon the level of consumption \( C \), but for simplicity it is often assumed to be independent of consumption, adopting an iso-elastic utility function, or constant elasticity of marginal consumption.

\[
U(C) = \begin{cases} 
C^{1-e} - 1 & \text{if } e \neq 1, \\
\log(C) & \text{if } e = 1
\end{cases}
\]

When \( e = 1 \), \( U(C) \) is the natural logarithm of \( C \), which suggest that a similar percentage increase in income across the income range leads to a similar enhancement of desire satisfaction.

Four main methods have been used to estimate \( e \). The first method uses consumer demand analysis applied to preference independent product groups (i.e. the utility of a commodity depends only upon the quantity of that commodity consumed) (Fellner, 1967), with a relatively small share of consumer budgets. Marginal utility of income is estimated by considering the responsiveness of demand to a change income relative to the responsiveness of demand to a change in price. A recent estimate using this method, considering the demand for food, found an estimate of 1.60 (Evans et al., 2005).

The second method estimates the elasticity of marginal utility of income from savings behaviour and lifetime consumption, on the assumption that households or individuals will aim to maintain a constant marginal utility of wealth. Blundell et al. (1994) use UK Family Expenditure Survey data (1970-1986) on income, consumption and interest rates to estimate a value of \( e \) from 1.2 to 1.4 and 0.35 to 1.05 when controlling for the high interest rates of the 1980s. Evans (2005) notes that given the financial shocks, high oil prices, and deregulation in
financial markets that occurred during this period, this figure may not be representative for the current UK. Hanna et al. (1995) list estimates from 13 studies published in the 1980s that show a wide range of estimates for the elasticity of marginal utility based on inter-temporal consumption decisions.

The third method uses surveys to gain a measure of risk aversion or inequality aversion. Amiel et al. (1999) finds an $e$ of between 0.2 to 0.8 from a measure of inequality aversion amongst students. Barsky et al. (1997) find a much higher value of 4.2 from direct experimental evidence on risk aversion in surveys of middle-aged respondents in the US. A recent estimate using this method was 1.63 (Evans et al., 2005).

Finally, the fourth method considers the degree of aversion to income inequality as revealed through preferences for the progressiveness of income tax – the income inequality aversion parameter. Evans (2005) finds a value of 1.4 from revealed social values from personal tax rates in 20 OECD countries.

Estimates within and between methods vary considerably. Pearce and Ulph (1995), in their survey of the evidence, give a range of estimates of $e$ from 0.7 to 1.5, with what they consider to be the best estimate, 0.83, coming from Blundell et al., (1994) and their analysis of saving behaviour. They conclude that a value of 1 is defensible and 0.8 to 0.9 would be relevant to current UK conditions (Pearce and Ulph, 1995: 14). This work contributed to the UK Treasury recommendation of a value of 1 to be used in economic evaluations (H.M. Treasuring, 2003).

Part of the disagreement between studies arises because in all of these methods diminishing marginal utility of income cannot be fully isolated from other effects. Measures of risk aversion may be picking up risk attitudes in addition to perceptions of diminishing marginal utility. Measures of aversion to inequality incorporate social preferences towards equity concerns in addition to efficiency measures on elasticity of marginal utility of income. And measures of saving behaviour will be influenced by the extent of an individual’s self-control, and depend upon assumptions about the value of personal discount rates.

An individual who is willing to incur considerable financial risk today is also likely to be prepared to risk the possibility of low consumption in the future because they may reason that in the future their financial situation may be altered or they may be dead. This suggests a strong relationship between risk aversion and inter-temporal consumption decisions. Likewise an individual with greater risk aversion in the present is also likely to favour constant consumption across their lifetime. However, these are still two different concepts. Aversion to social inequality is related to individual risk aversion through the individual’s preferences to possible
locations of their future self within any income distribution, but again it remains conceptually different to individual risk aversion.

Given the different methodologies and different concepts used, different studies are likely to generate different estimates for \( e \), and the correct choice of estimate will be open to debate. In the absence of definitive evidence, any rate chosen for use in CBA will be subject to potential challenge (Evans, 2005). In part as a consequence of this uncertainty, researchers and policy makers have generally failed to adequately identify and incorporate the elasticity of marginal utility of income into measures, and any adjustments to account for diminishing returns to income are treated as adjustments to account for equity concerns rather than efficiency concerns (Sen, 1979).

Although at a conceptual level there is little challenge to the idea of diminishing marginal utility of income, failure to identify the elasticity of marginal utility of income seriously undermines the ability to use income as a measure of value at both the aggregate and micro level and presents a serious gap for welfare economics.

1.3.2 Current use of welfare economics measures

It may be pointed out that the problems with welfare economics are well known. Even so, it continues to have a profound effect on government policy. Government policy covers an extremely broad range from protecting individual rights (of current and future members of society), correcting market failures and enhancing allocative efficiency, improving equity, and increasing well-being. In all but the protection of rights the use of income as a measure of value feeds into government decision making. Income as a measure of value is used to identify market failures, value most non market goods, establish levels of equity, and give an indication of overall social well-being.

The assumption that income (both national and individual) can be taken as a measure of well-being underlies much of the political emphasis given to income and economic growth. The development and importance of GDP accounts shows the dominant role played by maximising income in directing government policy. Whilst economic growth may be instrumental to other beneficial outcomes, the primacy of GDP is founded upon its interpretation as a measure of success of the society and an indication of the level of well-being of its people.

Since 1948 the standard System National Accounts (SNA) has been widely used internationally allowing both inter-temporal and international comparisons. Avner Offer describes the 1950s and 1960s as having a “competitive preoccupation with economic growth” in which GDP per
capita was taken as the “normative benchmark for economic and even social performance, the higher the better” (Offer, 2000: 4).

GDP growth has been a central, long-standing government policy, based on an assumption that this will increase individual income and hence individual well-being. However, measures adopted to maximise economic growth are rarely justified in terms of well-being, growth being taken as an end in itself. The status of economic growth is such that it no longer requires justification as an objective.

Similarly, at a micro level, WTP for goods or services is assumed to measure the anticipated change in utility or well-being following their acquisition. This is clearly seen in the use of WTP as a measure of value in economic evaluation methods recommended by the UK Treasury (HM Treasury, 2003).  

In order to make comparisons across a broad range of programmes and the myriad of benefits brought about by government intervention, from the reduced risk of death on the roads to the pleasure derived from public parks, it is necessary to express benefits in a single metric. Expressing benefits, such as environmental, social and health effects, in monetary terms enables full cost benefit analysis. This in turn aids decisions on both the relative benefit compared to other projects, and whether the projects are worthwhile.

Economists have developed methods for valuing benefits for which there are no market data. Some of these methods involve inferring a monetary value from observing consumer behaviour using ‘hedonic pricing’. Where the value of benefits cannot be inferred from ‘revealed preferences’, ‘stated preference’ methods have been developed, which elicit monetary values through hypothetical choices presented to the general public or particular population groups. For example, the contingent valuation (CV) method involves surveys in which respondents are asked to state their maximum WTP for a given benefit (Arrow et al., 1993). Within CV methods an individual is assumed to have stable indifference curves that represent their ordinal preferences between alternative goods. The difference between indifference curves may be measured by the equivalent variation in income or the compensating variation in income. The former is the money equivalent at the original prices that yields the same level of utility as the proposed change in non-monetary goods. The compensating variation is the money transfer at the new prices that returns the individual to her original level of utility. The compensating variation can be measured by an individual’s WTP for a positive change or his willingness to

---

3 Although the 2003 guidance clearly recommends the use of WTP for valuing market and non-market goods, including health losses and risk of mortality, supplementary valuation methods are also discussed and these guidelines adopt a more critical stance on stated preference methods than found in the previous, 1996, version.
accept (WTA) for a negative change and the equivalent variation can be measured by the individual’s WTA to sacrifice a positive change or WTP to avoid a negative change.

If each individual is given equal weight in a social welfare function, and if each individual’s marginal utility of income is assumed to be constant, then the simple aggregation of all individuals’ WTP or WTA can be used to represent the potential change in social welfare. Of course, if the utility function of all individuals is known, it would be possible to account for diminishing marginal returns to income. The UK Treasury Green Book recommends that CBA should ideally weight the costs and benefits of future projects by the estimated marginal utility of income for different income groups but, at a minimum, any distributional effects should be quantified as far as possible (HM Treasury, 2003).

Any macro policy or micro evaluation in which WTP, either revealed in the market or via stated preference, is used as a measure of benefit implicitly assumes that changes in income are a good reflection of changes in well-being. Although much has been written questioning the validity of WTP surveys, the fundamental link between changes in income and utility is usually assumed without question. If this link is not valid, then regardless of how accurately stated WTP could reflect the true WTP of a good/service were it to be available via the market, it would remain just that – a WTP for a good/service, rather than a reflection of the increase in well-being which that good would bring.

Despite this reliance upon income as a measure of well-being at both the micro and macro levels there has been general recognition that income is not the only outcome of importance within economics and other social sciences (e.g. Galbraith, 1958). Researchers, mostly psychologists and sociologists, have drawn attention to the lack of convergent validity of income in that it appears not to be related to other attributes that would be thought to reflect well-being.

1.3.3 Income as a measure of well-being appears to lack convergent validity

Government policies in the UK have successfully facilitated an increase in income as measured by GDP. If well-being is taken to be closely mirrored by income, they have therefore achieved success. This improvement is reflected in many other indicators which we would expect to correlate highly with well-being, such as increased life expectancy (Crafts, 2005), increases in leisure time, increased schooling, improved quality of living conditions, reduced corruption (Tanzi and Davoodi, 2000), reduced relative poverty, child poverty, child mortality and income inequality (Boarini et al., 2006).
Despite success in terms of income growth many other indicators of social success show reverse trends, for example, environmental damage and climate change, rising obesity, particularly amongst children (ONS, 2007), rising illicit drug dependence (ONS, 2000), a rising proportion of children living in lone parent households, and falling trust (Performance and Innovation Unit, 2002) and in the US rising anxiety and neuroticism amongst the young (Twenge, 2000). These are sufficient to doubt that well-being has risen in line with GDP.

One of the largest threats to the validity of income as a measure of well-being has come from the realisation that despite dramatic increases in income in the last 30 to 50 years in many Western societies self-reports of happiness have shown a static trend (Easterlin, 1974, 1995). To remain confident in income as a measure of well-being, we need to address why something which is expected to correlate highly with well-being does not appear to be correlated (over time) with income.

### 1.3.4 Problems with other measures

A range of adjusted income accounts have been developed to give a better reflection of the success of a society, many of which adjust for capital depreciation, unpaid household labour and regrettable necessities. For example Nordhaus and Tobin use an adjusted version of GNP for their ‘Measure of Economic Welfare’ (Nordhaus and Tobin, 1971). This includes estimates for the value of leisure time and some non-market production and deducts some costs that arguably do not contribute to economic welfare, such as commuting and public spending on police and defence.

More substantial adjustments are made to GNP in the ‘Index of Sustainable Economic Welfare’ (Daly and Cobb, 1989), which incorporates both environmental effects and changes in the distribution of income. However, such accounts are often criticised for lacking empirical and theoretical foundations, and, except among certain groups, they do not receive prominence.

The ‘social indicators movement’ (Sheldon and Freeman, 1970) of the 1960 and 1970s popularised the collection of data measuring quality of life. This incorporated collection of data on health, deprivation, housing and environmental conditions, amongst other variables (e.g. the UK Social Trends published since 1970).

Although social indicators are often highly correlated with income (for example, Diener and Diener (1995) found that the wealth of nations correlated 0.82 with number of books published per capita, 0.73 with income equality within nations, and 0.70 with the percentage of persons attending universities) they offer the potential to highlight important areas which may deviate from economic growth and provide a more detailed understanding of social processes.
One of the stumbling blocks for social indicators has been the problem of generating a summary index. For some researchers the multidimensionality of the concept of quality of life and incommensurability between dimensions implies that no simple ordered indicator of level of living can be constructed, either on an individual or aggregate level (e.g. Erikson, 1993).

However, the political desirability of a single indicator has led to a range of single quality of life indexes being developed. The most common weighting procedure is to use equal weights for all social indicators after each has been standardized. This includes the Human Development Index (UNDP), the Index of Social Health (Institute for Innovation in Social Policy), the Value Based Index of National Quality of Life (Diener, 1995), the Comprehensive Quality of Life Index (Johnston, 1988), and the Physical Quality of Life Index (Morris, 1979).

An alternative approach to weighting is the use of factor analysis which aims to maximize the variance explained in the social indicators, as used by the Weighted Index of Social Progress (Estes, 1997).

The choice of weights can result in a different picture being presented. This can be seen in the re-analysis of data from the 1995 Places Rated Almanac by Becker et al. (1987). This had quality of life data from 329 metropolitan areas of the US including variables such as climate mildness, health care, crime and recreation opportunities. Becker and colleagues found that, depending upon the weights given to the variables, there were 134 different cities that could be rated first and 150 different cities that could be rated last, and even 59 cities that could be rated either first or last.

Social indicators provide important inputs into policy. Two indicators that attract particular political and public attention in the UK are child poverty rates and unemployment. However, the absence of a well-established headline indicator of overall well-being limits the political role of social indicators. Furthermore, there has been no focus on measuring individual well-being within the ‘social indicators movement’ hence this has limited the understanding of distributional issues and has meant there is no outcome measure for micro level economic evaluation.

Aggregate income has been found to be inadequate both at the macro level and micro level as a measure of well-being. There have been attempts to address the limitations of income measures partly through adjustment of income measures, and partly through the development of additional social indicators. However, the political weight given to GDP growth suggests it is still seen as both a contributor to quality of life and an end in itself to be traded against quality of life.
1.3.5 The current state of measures of well-being

The main measures of well-being used within policy are still drawn from welfare economics. As discussed above there are many theoretical weaknesses within welfare economics, and the validity of the assumptions attracts a growing body of criticism.

Other measures have their own theoretical limitations that are in part responsible for their failure to be adopted on a large scale, or challenge the predominant use of income as a measure of well-being.

1.4 Reconsider measures of well-being

It is necessary and timely to reconsider the emphasis given to income as a measure of well-being and compare the pro’s and con’s of income versus other alternative measures of well-being due to a number of reasons.

Firstly, as a result of increasing awareness of the shortcomings of the use of existing measures, government bodies, particularly in the UK, are calling for new measures of well-being. Secondly, new measures based on subjective accounts of well-being are increasingly available, including evidence on their psychometric properties, and these are being advocated by psychologists for use in public policy. These shall be considered in detail below.

1.4.1 The UK Government is calling for a new measure of well-being

The objective of economic growth and maximising individual income is supplemented within government with many other well-being related objectives, clearly reflecting a growing understanding that income is an incomplete measure of current and future well-being. For example, in the UK well-being measures have recently been openly incorporated into monitoring and policy making. Since the 1998 Comprehensive Spending Review each major department has a Public Service Agreement (PSA), setting out department objectives and targets for meeting these and many are well-being focused.

The UK Government clearly has concerns over the use of income as a measure of well-being, as seen for example, in the Prime Minister’s Strategy Unit paper on the implication of life satisfaction research for government policy (Donovan and Halpern, 2002). The UK Government’s 2005 sustainable development strategy, ‘Securing the Future’, implies that well-being is an important aim of government which requires appropriate indicators to ensure its accurate measurement. As the report notes, “the issue of well-being lies at the heart of
sustainable development, and it remains important to develop appropriate well-being indicators” (HM Government, 2005: 23).

Indeed, steps to establish a new measure of well-being have already begun in the UK. In 2006 Defra commissioned preparatory work on the links between well-being and sustainability (Dolan et al., 2006b, Marks et al., 2006) and the different conceptions of well-being and their determinants (Dolan et al., 2006a), in order to feed into future development of well-being indicators. In 2007 Defra provisionally introduced a range of SWB indicators into its sustainability indicators, with a view to extending these in the future (Defra, 2007).

The lack of a clear definition of the terms ‘quality of life’ and ‘well-being’ have been claimed in a report by the Scottish Executive to lead to a “‘back to front’ approach to enhancing quality of life and well-being in communities and for individuals: what can be measured is measured, and what is measured drives the policy” (Scottish Executive Social Research, 2005: Chapter 2.3). This suggests a need to provide policy makers with valid measurement of individual well-being.

1.4.2 New measures of well-being from psychology

Psychologists have been using measures of SWB for many years. Recently, however, they are becoming increasingly vocal about the need to incorporate SWB within policy (e.g. Kahneman et al., 2004a, Diener and Seligman, 2004). However, the recommended measures range from global evaluative measures to moment-to-moment feeling based measures. For example, the following two quotes by key SWB researchers present quite different measures.

“Global measures of subjective well-being, such as assessments of life satisfaction and happiness, can be useful for policy debates” (Diener, 2006: 398).

“…we propose measuring national well-being by weighting the time allocated to various activities by the subjective experiences associated with those activities” (Kahneman et al., 2004a: 433).

This thesis aims to show that subjective measures of well-being offer great potential for policy, suggesting a need to move away from a preference satisfaction account of well-being within economics. However, such a dramatic shift is unlikely to occur unless alternative approaches to measuring well-being have theoretical rigor and clear practical advantages over current methods. There may be good reasons for favouring each of methods referred to in the quotes above, it is however, essential to understand the theoretical construct which supports each measurement instrument, and the strengths and weakness of each approach.
1.4.3 Summary of the need to reconsider measures of well-being

Inadequacies with existing well-being measures make it timely to reconsider which measure most closely meets the needs of policy makers. From a policy perspective this assessment should have the following features: i) each measure should be assessed according to the same criteria (which I establish in Chapter Four); ii) theoretical and empirical criticisms should be recast in terms of their practical implications; and iii) the superiority of any measure must be clearly shown.

For researchers in the area of happiness and SWB these are particularly exciting times in which the scope and coverage of research is expanding at a dramatic rate. This is also a critical time for clarifying the theoretical foundations behind many of the measures of well-being used in this literature. A lack of conceptual clarity at this stage is likely to have negative consequences on the potential of this movement to address policy concerns or to be incorporated within the economics discipline in the longer term.

1.5 What type of measure of well-being is required?

There are three levels of well-being that policy makers need to know: current well-being, well-being over a fixed duration of time, and lifetime well-being. Without a measure of well-being over time it is impossible to consider the net effect of actions that may have short-term costs for long term future gains, and without a measure of lifetime well-being it is impossible to assess policies that have implications for both length of life and quality of life.

If higher momentary well-being always resulted in a proportionally higher long-term or lifetime well-being then for policy purposes knowledge of current well-being would be sufficient. However, there may be conflicts and trade-offs between current well-being and long-term well-being. For example, an individual may sacrifice current well-being in order to reap rewards at a later date (e.g. sacrificing eating unhealthy food, studying for qualifications, stepping outside in the cold for a long run) or similarly may increase their current well-being at the expense of future well-being (e.g. having the third glass of wine).

The importance of understanding the time scale involved within measurement has been given little emphasis. Whilst it is not as necessary for understanding what is of value, it becomes essential if we wish to use well-being measures to evaluate policy.

Measurement requires identifying the constituents of well-being, or what comprises an individual’s well-being, and a means of accurately representing those constituents. It is also desirable to have a clear, defendable theory of why those constituents comprise well, or what
makes the constituents of well-being good for people. Without such clarity (if say relying only on historical precedent) the measurement instrument will be vulnerable to challenges.

1.6 Outline of the thesis, methodology and contribution

This thesis draws upon work on well-being from philosophy, welfare economics, health economics, sociology, psychology, and recent empirical work exploring the relationships between well-being and personal, economic, and social factors. Each section contains a revision of the relevant literature.

Issues requiring additional evidence will be supplemented, where possible, by new evidence drawn from analysis of the British Household Panel Survey (BHPS). This dataset offers detailed information on an unbalanced panel of initially 5,500 families and 10,000 individuals in 1991 expanding to about 15,500 individuals in 2005 (details can be found in Appendix I). The BHPS contains a wealth of information on personal characteristics, income, expenditure, household demographics, health state, and subjective indicators of well-being. As such it is ideal for exploring alternative measures of well-being.

Chapter Two begins by asking exactly what is meant by individual well-being. It is not possible to begin looking at measurement issues without first establishing clarity in the concept of well-being. The analysis draws from philosophy, where this question has been a topic of debate for over 2,000 years.

A brief overview of the main philosophical accounts of well-being is given, summarising the variations within each account, considering what they hold well-being to be, and why this is taken to be well-being. The aim is not to find a better conception of well-being, nor raise new objections to existing theories, rather the focus will be on measurement. Theories of well-being draw upon moral philosophy and are not generally intended for the purpose of measurement, but rather for understanding what is of value. However, in order for policy to be influenced by well-being some form of measurement will be necessary. Some accounts lend themselves better to measurement than others. Not being amenable to measurement is not seen as a critique of the account itself, but weakens its relevance for policy making.

Although much has been written about well-being, little work has been done on translating concepts of well-being into measurement instruments, and very little which aims to incorporate a time dimension into measurement. However, a clear time dimension is essential for policy. As shall be seen, incorporating a time dimension requires additional assumptions. A discussion of these is given.
Much work on measuring well-being tends to focus within a discipline. I draw from a range of disciplines and discuss existing measures that have been used in practice. Options for measurement are developed through considering the steps and assumptions necessary to take the theoretical conceptions of well-being towards practical measurement.

This thesis aims to consider empirical differences between individual measures of well-being and give a systematic assessment of different measurement approaches. Both endeavours have as a prerequisite a clear conceptual framework for understanding differences between measurement instruments and for identifying the strengths and weakness of both measurement approaches and specific instruments.

We need to base empirical work on theoretical understanding of the concepts employed. The concept of ‘utility’ has remained somewhat nebulous within economics yet if we wish either to reject or support new measures of well-being understanding the differences between measures is essential. Chapter two shows that preference satisfaction and mental state accounts of well-being embody very different approaches to thinking about well-being.

For each theoretical approach one or two examples of measurement instruments are given to demonstrate the type of measure that is compatible with each account. Instruments chosen are: those which are present within the BHPS (these form the basis of empirical work in Chapter Three); those which are used within UK policy area; and those which are commonly included in large scale social science surveys which have been frequently used within the SWB literature.

Understanding the theoretical differences between accounts of well-being helps explain what currently employed measures of well-being might be identifying and when they are likely to be appropriate for use in policy.

Chapter Three aims to explore the empirical differences between well-being measures, both between different instruments within the same conception of well-being and between measures using different conceptions of well-being. I consider the extent to which an individual’s well-being differs depending on the choice of measurement instrument. The aim being to give some context and potential practical relevance to theoretical differences raised in Chapter Two. If two measures appear theoretically very different, yet in practice they lead to similar judgements about individual well-being, and to similar policy, then the relevance of theoretical differences should not be overstated. Analysis of the impact of the choice of well-being measure is made by considering its relationship to personal, economic, and social factors. Unfortunately, comparisons across studies are often of limited value due to the many other factors that differ between studies in addition to the well-being measure. Furthermore, there are an insufficient number of studies that contain more than one well-being measure. Consequently, exploring the
differences between well-being measures will be supplemented by analysis of the BHPS, which is ideal for making such comparisons. The comparisons I make draw upon ten different measures of well-being, a far broader range than currently exists.

I identify those subgroups for whom choice of well-being measure matters most. This analysis shows that, in many cases, choice between different SWB measures appears to make minimal difference, although there are some important cases where SWB measures show a different relationship to other factors such as gender and education. In the case of gender these differences have been given reasonable explanations. However, the case of education presents more of a puzzle. Some evaluative style SWB measures show a positive relationship to education, others show a negative relationship. This is explored through a novel consideration of the response scale of SWB measures and the use of statistical analysis that allows relationships to vary at different points in the response scale.

More notable (and policy relevant) differences are found when comparing SWB measures as a whole with those of income and consumption.

Choosing between different measurement options will be easier when the decision is based upon a clear, systematic set of criteria. The approach of setting out clear criteria by which outcome measures can be judged is well used within health care research as a means of supporting high quality new research and judging the quality of existing research. The National Institute of Health and Clinical Excellence (NICE), for example, uses a set of criteria to determine suitable health-related utility measures (NICE, 2004).

There are no existing criteria for measuring well-being for public policy. I formulate a set of criteria drawing from the literature, the discussion in Chapters Two and Three, psychology and psychometrics, and practical and policy considerations determining whether a measure would be successfully implemented.

Chapter Five considers the preference based and mental state accounts of well-being and asks how the measures available within each account fare according to the criteria outlined in Chapter Four. This highlights the weaknesses inherent in the different conceptions of well-being and different instruments available. This methodology produces a resource for policy makers when addressing inevitable disagreement over measuring well-being.

One of the areas in which income is assessed as inadequate is in terms of validity (i.e. measures what it claims to measure). In order for the measure to meet the criteria of validity it is necessary that the measure provides sufficient information about an individual’s well-being. In the case of income, this would require that current income or consumption is the only factor
within an individual’s utility function. However, a large body of literature has raised the possibility of interdependent utility functions (Pollak, 1976). The happiness-economics literature has made a substantial contribution to the understanding of how previous incomes and other people’s incomes impact upon utility, further undermining validity of income as a measure of well-being.

Chapter Five finds that SWB measures fare relatively well according to the criteria mentioned above. Consequently, the next three chapters consider how SWB can be operationalised within a policy context, and whether findings from the SWB literature can be incorporated into policy making.

Chapter Six explores policy issues concerning the relationship between income and SWB. Three implications drawn from the SWB literature are explored: i) unanticipated adaptation to income implies inefficient trade-offs between income and other goods; ii) relative income effects lead to collective overwork and inefficient trade-offs between income and other goods and iii) relative income effects lead to over-consumption.

The existing literature is considered for each implication, and additional analysis is conducted on the BHPS to illuminate the processes involved, and to consider whether empirical evidence offers support for the suggested policy conclusions. The relationship between income and SWB is crucial to policy. A good understanding of this relationship can highlight problems in equating well-being with income, and contribute to improved fiscal policy and overall government priorities.

My analysis on the BHPS casts doubt on the understanding of a simple annual adaptation to income. Although identifying a negative impact of past income on SWB, when controlling for current income, additional tests on this relationship suggest that it is being driven by processes that are yet to be fully understood. Negative relative income effects are often cited, however, consideration of the literature suggests conflicting impacts of others income. My analysis suggests the negative impact may be confined in the BHPS to those who move geographical areas. The third policy relevant implication, that others income may lead to over-consumption, has no support within the BHPS. In each case a clear theoretical reason for the potential policy recommendation is shown, however, analysis on the BHPS suggests the need for more empirical evidence to support any policy recommendations.

Chapter Seven considers an example of a determinant of SWB other than income. The analysis in Chapter Three showed that many factors associated with increased SWB are associated with increased income. For example, in situations of unemployment, poor health, and marital
Chapter 1 Introduction

separation individuals fare poorly in terms of both income and SWB. Consequently, a switch to focusing on subjective measures would be unlikely to lead to any dramatic policy shifts. However, the relationship between social capital and income on the one hand and social capital and SWB on the other diverges. As also shown in Chapter Three, social capital has a negative relationship with income yet positive relationship with SWB. A shift to SWB outcome measures may result in policy targeted at enhancing social capital that would not arise when income is used as the outcome measure.

Chapter Seven explores whether the understanding gained from analysis of large datasets on the correlation between social capital and SWB can be useful for policy. This analysis specifically focuses on one aspect of social capital, the relationship between talking to neighbours and its relationship to SWB. Through analysis of the BHPS I consider whether this relationship is causal, since evidence showing causality is essential to support policy recommendations. Although similar analysis has been conducted (Powdthavee, 2005, 2007), this is useful because the focus here is upon different variables, and uses different instruments to address potential endogeneity of social capital variables. Furthermore, talking to neighbours represents an ideal case study in which to explore the success of using instrumental variables to address the issue of endogeneity, and offers support for one of the main recommendations of the thesis – the need for experimental and quasi-experimental studies.

Chapter Eight moves into the micro-level and explores the use of SWB measures as a metric of value, both directly and as a means of establishing monetary valuations for non-market goods. This will consider the strengths and weakness of valuations based on SWB data in comparison to the more conventional use of WTP. Since monetary valuations generated from SWB data are easily estimated, it is important to be clear about their contribution to policy and exactly what this information is telling us. Although income compensations from SWB are increasingly used, little has been written about the limitations of this method from a policy perspective, nor its relative advantages over WTP. Concerns raised over the use of monetary compensation using SWB data will be explored through examples. I argue that given the problems with generating compensation-income through SWB models, a more constructive approach is to focus only on absolute changes in SWB.

Chapter Nine concludes by considering policy recommendations that can be drawn from this research. Recommendations are also made relating to extending this work, both in relation to public policy and research. More research is recommended into exploring the extent to which an assumption of interpersonal comparability and cardinality is appropriate for SWB measures. This requires a greater understanding of the processes individuals use when responding to SWB survey questions. New research is proposed to enable SWB measures to be incorporated into
economic evaluation. The knowledge gaps required to extend the research on the determinants of SWB into a policy agenda are identified. Drawing from the Chapters Two to Five recommendations are suggested for the most appropriate current means of measuring well-being for the UK.

The overall contribution of this work rests in part on its analytical synthesis of literature relevant to measuring well-being, but more importantly on the multi-disciplinary approach adopted. I hope to show that empirical analysis should be grounded in an understanding of what we mean by individual well-being and what we expect our measures to identify. If we care about resource allocation decisions, which we should, we need to provide policy makers ways of monitoring individual well-being and evidence-based policy recommendations. Through a systematic assessment of well-being measures, a strong argument for reducing the emphasis given to income and WTP studies in favour of SWB measures emerges. SWB measures are not presented as the perfect policy tool. However, even in their current form – warts and all – their incorporation into policy offers a means for policy makers to improve well-being.
Chapter Two: Measuring well-being

2.1 Introduction

Policy makers seeking to enhance individual well-being are faced with a variety of different instruments proposing to measure it. For example, the Australian Centre on Quality of Life lists over 900 scales that aim to measure all or part of individual well-being.\(^4\) Differences between these measures arise, in part, because of differences in the understanding of well-being itself.

Accounts of well-being can refer to two quite different concepts: a life which is good for the individual, and the ‘good life’. The former is equated with individual prudential value, whereas the latter can incorporate other values, such as aesthetic or moral values. Chapter Four will discuss which of these is most suitable as a measure of well-being for public policy.

Furthermore, there are different ways of thinking about well-being within those two overall conceptions. Consequently, an understanding of measurement issues must first address the different ways of describing and conceptualising well-being.

To understand what is meant by well-being we turn to the well-developed dialogue in the philosophy of well-being. Theories fall into three main categories: preference satisfaction accounts; objective list or substantive good accounts; and mental state accounts.\(^5\) In preference satisfaction accounts an individual’s well-being is judged by the extent to which her preferences or desires are met. Substantive good accounts make well-being depend upon success across a range of specified attainments, such as health and literacy. Mental state accounts link an individual’s well-being to her cognitive and emotional responses to her life.

Measurement and consequent empirical testing of substantive good accounts has been limited, in part due to their structure in which goods are perceived as non-commensurable hence not lending itself to a single well-being scale. Furthermore, there is no available data within the BHPS which is compatible with any measures drawn from a substantive good perspective. For this reason, substantive good accounts are not included in this chapter, nor in the assessment of well-being measures in Chapter Five. In order to present a comprehensive review of measures of well-being, they are given the same treatment within Appendix III.


\(^5\) The tripartite categorisation of Parfit (1984) is commonly used in the literature (e.g. Griffin, 1986; Crisp, 2001). However, Scanlon’s (1993) term ‘substantive good conceptions’ is used here instead of Parfit’s ‘objective list’ accounts as many of what could be described as objective accounts contain elements of subjectivity, hence ‘substantive good’ gives rise to less potential confusion.
Chapter 2

Measuring well-being

The first aim of this chapter is to set out each theory. The second aim is to consider the means by which well-being according to each conception could theoretically be measured, in the moment, over a specified duration of time and over a lifetime. This will be supplemented with examples of compatible measurement instruments. These are chosen to indicate the range of measurement options available for each account. In many cases the chosen instrument represents a number of possible instruments which adopt a similar structure. Incorporating all measurement instruments found within health, psychology, economics and sociology which claim to capture aspects of individual well-being would distract from the central purpose of this work. However, the example instruments need to be sufficiently representative to highlight issues arising at the empirical level. The instruments chosen here are those which are found within the BHPS, since they will be the focus on empirical work in Chapter Three, those which are most commonly referred to within the well-being literature, and those which have been raised at UK policy levels.

2.2 Preference accounts

2.2.1 Theories of preference or desire satisfaction

The first group of well-being accounts are preference satisfaction accounts. Well-being is seen by preference satisfaction or desire fulfilment theorists as a state of the world in which a person has achieved what they desire. Preference satisfaction theories fall into three sub-categories; actual desires, restricted desires and idealised desires.

At the simplest level, all desires which an individual actually holds count towards his well-being, hence, “what is best for someone is what would best fulfil all of his desires, throughout his life” (Parfit, 1984: 494). Thus, people are well off to the degree that their actual preferences are satisfied.

This simple account is open to an obvious objection, that there are many examples of desires, the satisfaction of which do not, intuitively, enhance an individual’s well-being. For example, satisfying the desire to own a Flowbee Vacuum Haircut System, following late night indulgence in infomercials (Heathwood, 2005: 488) does not appear to raise well-being.

In response to objections of this kind, theorists have either restricted the type of desires which should count towards well-being, leaving only a subset of relevant desires, or presented well-being as the satisfaction of idealised, hypothetical desires.

---

6 How have I coped without one? - http://www.flowbee.com/
Theorists who restrict the type of desires look for a feature in the desires to warrant their inclusion (e.g. relating to life goals (Keller, 2004: 33) or referring to one’s own life (Parfit, 1984)), or justify their exclusion (e.g. pointless (Rawls, 1971: 432), artificially aroused (Heathwood, 2005: 488) or intransient (Tefler, 1980) desires). The reason these features are excluding is that they can be thought of as faulty or irrational. Pointless, unsatisfying, and time inconsistent desires may be thought of as faulty or irrational as well as irrelevant to how well an individual’s life is going. Faulty desires include cases where the individual does not have full information about an object of desire (such as wanting to eat an unknowingly poisoned apple) or desires which are not stable under experience and reflection (Gauthier, 1986: 32-33, cited in Moore, 2004: 77).

Hypothetical, idealised preferences are those preferences that an individual would have held had certain conditions, considered ideal for forming preferences, been met. John Harsanyi describes those ideal conditions as those in which an individual “had all the relevant factual information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice” (Harsanyi, 1982: 55).

Whilst restricting and idealising desires overcomes many of the potential objections to well-being as desire satisfaction, they raise additional concerns about how the selection criteria are to be established. The more restricted and idealised desires become, the further removed the account becomes from one based on the premise that an individual’s well-being is the extent to which individual gets what he wants.

2.2.2 Measuring desire satisfaction

Measuring preference satisfaction can adopt one of two approaches; an aggregation of the attainment of the different desires or a position on a preference ordering of possible lives. Both these approaches shall be considered.

Aggregating preferences

Measurement issues raised by the aggregating preferences approach

Measurement according to the first approach, summarising the extent of desires met, raises some difficult questions.

How do we measure how well or completely a desire has been met? To do so may require knowledge of the individual’s mental response to the situation, or knowledge of the state of the world beyond their own experience.
How do we measure the intensity of a desire, or the importance to the individual that the desire is met? To generate a summary measure requires some kind of weighting system such that success of some important desires, such as wanting a good relationship with one’s parents, is given higher value than success in other smaller desires, such as wanting a cup of good coffee. This may be, for example, based on the felt intensity of each desire. However, judging strength of desire by felt intensity has been criticised by Griffin (1986) since this may be a product of upbringing (emotional versus stiff upper lip). Judgement in terms of motivational force is also problematic since one may wish for something that contravened second order desires.\(^7\)

How do we generate an overall well-being level from the various met and unmet desires? Should we take well-being to be the summation of all desires met, the net of desires met less desires not met, or the percentage of total desires met over all desires held? A simple summation of each desire met means that behaviour or influences that generate new desires which are met leads to an improvement in well-being, even if those behaviours and influences also lead to many more desires which were not met. Parfit raises an objection to the simple summation of desires met via the case of addiction, where the addictive nature of a drug creates an extremely strong desire, despite the fact that it brings no pleasure or even causes suffering. If well-being is an aggregate of all desires met weighted according to their strength, then a strongly desired, cost-free addictive drug would theoretically enhance well-being (Parfit, 1984: 497).

Measuring well-being over a specified duration of time, within this summation framework, faces further complexities. For example, should a met (unmet) desire be included in every period when it remains met (unmet)?

To generate well-being over a specified duration of time, momentary well-being may be aggregated. Broome argues that the most reasonable default position is an equally weighted aggregate of moments (Broome, 2006: 223). He also shows that such aggregation is only possible if moments and people can be treated separately. Hence it requires that each moment in a person’s life must be evaluated independently of how things go in other moments or for other people.

Consequently, this approach ignores preferences regarding the order and path of experiences. For example, an individual may prefer to have an improving rather than deteriorating life, or give particular weight to peak and end moments, or value evenness of outcomes across their life (Broome, 2006).

---

\(^7\) Second or higher order desires are desires to have certain desires or wants such as the desire not to want to eat meat if one is a vegetarian.
David Velleman (1991) argues that momentary well-being and well-being over an extended period are irreducible elements of welfare. He argues that the story of a life is better if it improves rather than deteriorates, therefore, any simple aggregate of moments will not capture the important narrative quality of a life. Consequently, aggregation over time is problematic for the summation of desires as an account of well-being, since from the individual’s perspective the value of a profile of experiences depends not just upon the moment-to-moment experience but attitudes towards the profile itself.

Measuring well-being over a lifetime raises yet more complexities. If well-being is the total number of desires satisfied, adding additional years of life will increase total well-being since at least some desires must be met for an individual to be alive. However, using the net of desires met to desires not met, or percentage of met desires could mean that an additional year of life in which many desires are not met may reduce lifetime well-being. This is a similar problem to the average versus total well-being debate at a population level (see Hudson, 1987).

A further complication for measurement according to the aggregation method arises if an individual’s preferences change during their lifetime. This generates the potential for desires held and satisfied in the past to be perceived as detrimental to well-being judged from the perspective of the future. For example, the desire for alcohol met yesterday may be regretted today. In absolute preference satisfaction, the most reasonable approach would seem to be to weight each desire by the strength of the desire as experienced in the moment. Therefore satisfying the desire to drink alcohol increases well-being by the strength of desire for the drink, and the desire not to have drunk alcohol the following day decreases well-being by the strength and duration of that conflicting desire. This example shows that some desires are revealed through action (the purchase of a drink), other desires, however, are revealed by both non-action and the regret of action.

The restricted and idealised desire accounts appear to have an answer to this time inconsistency problem. If desires are restricted, then we may include the strength of only certain desires, such as our desires which contribute to our lifetime goals. Whilst these restricted desires may still change over time, they are less likely to be time-inconsistent. Similarly, if desires are idealised in some way, then a measure of momentary well-being will be an external assessment of how the individual’s life would meet certain hypothetical desires. If the preference structure itself is fully idealised and in line with what is important for actualising human capacities, then there is no reason to believe preferences would change.

However, measuring the summation of desires from an idealised desire satisfaction perspective is problematic because those desires remain hypothetical. One option may be to create an
environment in which the individual is encouraged to reflect upon their highest order
preferences, to provide them with sufficient information about their life (e.g. their objective
health state, predicted life expectancy etc.) and to ask them, from that reflective position, to
judge the extent to which their idealised desires are satisfied. However, this would still require
the assumption that their reflective, higher order desires were in line with idealised desires.

Another option is to assume that an individual’s idealised, self-referencing preferences could be
inferred by an outsider, on the basis that if any human being carefully, and with full
information, considered what they want out of life for themselves, the common nature of the
human species should lead them to identical conclusions. This is the approach adopted by
Griffin (1986) who argues that a fully informed desire account of well-being will equate with an
objective list account.

Measurement instruments for the aggregation approach

Possible measurement instruments which tap into the summation of desires are ‘gap’ theories
which draw from health care research in the 1980s. These present quality of life (QoL) (which is
this case appears synonymous with well-being) as the divergence or gap between an
individual’s life circumstances and the standard to which their life is compared. For example,
Alex Michalos’s ‘multiple discrepancy theory’ views QoL as the gap between an individual’s
expectations and his experiences (Michalos, 1985). QoL is defined by Camfield and colleagues
as, “the outcome of the gap between people’s goals and perceived resources, in the context of
their environment, culture, values and experiences” (Camfield et al., 2005).

If an individual assesses the gap between their life and their preferred life this would seem to be
a general measure of the extent of actual desire satisfaction, where extent is judged by the
number of unmet desires. Alternatively, wants and needs may be determined externally, based
upon inherent human needs, in which case this would be a measure of the extent of idealised
desire satisfaction, where extent is judged on the basis of unmet, idealised desires. The wide
range of comparison possibilities within gap theories (from a specific comparison group, to
what one considers ideal, and what one has desires for (Hass, 1999)) means that measurement
derived from ‘gap’ theories may be more closely aligned to different theoretical accounts of
well-being, depending upon the comparison group chosen. If current attainment and
expectations are both idealised then this perspective could be compatible with either an
idealised desire satisfaction account or a substantive good account of well-being.
Whilst measurement based upon gap theories is possible,\(^8\) using such measures for well-being over a specified duration of time is problematic since the measure invites changes in expectations to influence the gap. Using gap theories to measure changes in both quality and quantity of life is problematic, since although repeated measures may pick up a profile of well-being, there is no trade-off inferable in the measurement between additional years of life against changes in the gap between expectations and achievements.

Another option for measuring desire satisfaction according to some type of summation of desires met is to ask people the extent to which they have had their desires met or got the things they want from life (e.g. Bradburn, 1969). Self-reports can, of course, only include those desires that the individual knows have been met.

As discussed in Chapter One, the best measure of individual well-being arising from welfare economics is income/consumption. Current consumption (in the broad sense of all consumption not just market consumption) is most closely aligned to the extent of current desires met. This could, at least in theory, be measured directly and would include consumption of all goods that are of value to the individual during a particular time period. This includes consumption of purchased goods (e.g. food), on-going consumption of goods purchased in the current and previous period (e.g. dishwasher), consumption of non-market goods (e.g. security, nature), consumption of publicly provided goods (e.g. education, street lights), consumption of leisure and consumption of bads or regrettables (e.g. commuting to work).

It may be possible to indirectly measure consumption through current net income plus changes in assets, and accounting for the types of consumption noted above.

In order to include those things of value for which an individual has no opportunity to express their preferences within the market, a measure of the strength of desire for these states/items is required. If willingness to pay (WTP) is taken as an indication of strength of desire, and a hypothetical WTP is taken as equivalent to actual WTP, then hypothetical WTP values can be taken as the strength of preferences towards these states/items. For example, WTP may be used to incorporate preferences towards security, public services and the environment. However, whilst WTP is used to inform policy decisions, the range of states/items for which WTP figures exist does not cover the range of desires for which individuals have limited opportunity to express their preferences via the market. Consequently, measures of consumption tend to focus on the consumption of market goods. The extent to which this represents complete current

---

\(^8\) see the Bath Well-being in Developing countries group measure of individual quality of life http://www.welldev.org.uk/research/methods-toolbox/qol-toolbox.htm.
consumption depends upon how valuable non-market goods and bads are relative to market goods and bads.

Well-being over a specified duration of time (or over a whole lifetime) may be proxied by total consumption over that period. Lifetime consumption will equate with aggregate consumption (correctly measured) at each moment in time. Lifetime market consumption will equate to lifetime income (as long as bequests are included in consumption, and inheritance within income). Hence this model’s measurement approach appears consistent across time frames.

In order to treat aggregated current consumption as a measure of the satisfaction of desires over a period of time, certain conditions must hold. Desires themselves must be independent of consumption, the order of consumption must not matter and the rate with which consumption leads to desire fulfilment (the marginal utility of consumption) must be constant.

Aggregating current consumption to represent total individual well-being also assumes that total well-being is an equally weighted aggregation of momentary well-being. The required assumption that moments and individuals can be treated separately is highly problematic for income and consumption measures, and will be discussed in more detail in Chapter Five and Six.

Decisions must be made about precisely what type of income measure to adopt to represent preference satisfaction. Most income reports are of gross income but net income more accurately represents an individual’s choice set. There are also important issues surrounding the use of household or individual income, and how best to allocate household income across members.

The link between household income and desire satisfaction for the individual depends on the household composition. The more individuals that share any given income the less resources may be commanded by each individual. At the simplest level per capita household income can be used, however, this does not address two important issues. Firstly, how best to allocate household income to the different household members, allowing for age and potential power imbalances (Haddad and Kanbur, 1990). Secondly, how best to account for economies of scale in consumption. Household income can be ‘equivalised’ to account for the size of the household and the age of its members thereby giving a better indication of the desires that an individual is able to meet.

A range of different equivalence scales exist (Atkinson et al., 1995). The use of different equivalence scales can produce different results, particularly in relation to income inequality (Atkinson et al., 1995; Coulter et al., 1992; Banks and Johnson, 1994), suggesting that the
choice of equivalence scale influences the level well-being which would be attributed to the individual.

It can be argued that people choose the number of people living in a household, either in terms of current consumption or future consumption in the form of an investment (Pollak and Wales, 1979). If an individual chooses to have an additional household member, and this is a rational choice, the benefits (such as satisfaction of desires for companionship or future income) should equal or exceed any alternative benefits derived from the financial resources used by the additional member, in which case non-equalised household income would be a closer representation of total desire satisfaction. The assumption of rational choice appears more reasonable in some situations (such as a couple deciding to have an additional child) than others (such as a sick parent moving into the household, or unplanned pregnancy).

If an individual chooses to acquire a net-cost cohabitant then the utility gained from the cohabitant is assumed within welfare economics to be at least as great as the cost incurred. Whilst this can be argued for the individual who is a net contributor to household finances, using individual income for the individual who is the net gainer in the household seems less reasonable. For example, a non-working spouse may have an individual income of zero, but have equal access to their partner’s income. Consequently, in this set up one partner sacrifices income for the pleasure of living with their partner, yet the non-working partner has to be compensated. The root of this problem lies partly in the failure of income to incorporate non-market labour and more generally in the inherent inconsistency between seeking a measure of income that most closely represents individual well-being, whilst simultaneously acknowledging that income only represents part of individual well-being.

A commitment to preferences as the source of value would suggest that an individual’s preference over the combination of different incomes and family/household composition would be the ideal measure. This takes us back to the second method for measuring the extent of preference satisfaction, that of preferences for different lives given the individual circumstances of each separate person living that life.

Preferences towards whole lives

Measurement issues raised by preferences towards whole lives approach

James Mirrlees has argued that this approach of preferences between different selves in different states is the only way in which “measurability of utility can be achieved” (Mirrlees, 1982: 65). Alternative lives in different circumstances are assumed to be identical except for differences
that can be described. An assumption can then be made that an individual’s preferences for different lives will reflect the level of utility (or well-being) present in each life.

The individual’s current life may be thought of as a position in a preference ranking of all possible lives the individual can imagine, or a position on a universal scale of all possible lives (Weintraub, 1996), in line with absolute and idealised preference satisfaction. The latter suggests individuals share the same preference ordering between different lives. The former, however, raises two problems, firstly that an individual’s perceived position will depend on their imagination of the best and worst life possible; and secondly, that inconsistencies may arise if an individual’s preferences change over their life time. The former is discussed in detail in Chapter Five, the latter is discussed below.

Judging the utility attainable by different selves is a judgement from one particular vantage point. Unavoidably that vantage point will be one possible self with a particular set of preferences. As clearly put by Broome, “If a person has different utility functions at different times in her life - and she normally will - it cannot plausibly be claimed that any one of them represents her good” (Broome, 1985: 24). For the purposes of measurement, it must be assumed, therefore, that the individual will hold those same preferences in all lives, regardless of the different circumstances of those lives, otherwise no unique ordering of lives can be attained.

In Arrow’s version of ‘extended sympathy’ preference orderings between individual i in state x compared to individual j in state y, should be the same across all individuals as long as everything which determines the individual’s satisfaction is included in the list of goods within each state (hence the ability to enjoy wine as well as having the wine) (Arrow, 1977). It is of course questionable whether meaningful preferences towards states can be determined while asking the individual to imagine experiencing another set of preferences. This raises interesting issues relating to inter-temporal and interpersonal comparisons which will be discussed in more detail in Chapter Five.

Measuring an individual’s preferences for different lives is practically difficult because of the difficulty of classifying lives into a manageable number of identifiable circumstances. However, preferences demonstrated towards different health states, developed by health economists, share similarities with this approach. These use the individual’s current preferences towards possible future lives in different health states.

*Measurement instruments based on preferences towards whole lives*
Preferences towards the type of life being lived may be valued in terms of willingness to sacrifice future years of life in order to live in a particular state. Quality adjusted life years (QALYs) were developed as a means of valuing different health states in this way (Drummond et al., 1997). An individual’s preferences towards a particular state are expressed on a scale between 0 (for death) and 1 (for full health). This preference weight can be derived in three main ways: by asking the individual to place a health state on a visual analogue scale (VAS) anchored by 0 (death) and 1 (full health); by asking individuals to give the point of indifference between remaining in that state and accepting a risk of full health or death (standard gamble (SG) method); or by asking the individual to give the length of time in full health that they would consider equivalent to a given length of time in the health state in question (time trade-off (TTO) method) (Dolan, 2000). Hence as long as certain assumptions hold, cardinal data on the level of utility of health states can be attained.

Various preference based QALY instruments have been developed which use multi-attribute health-state descriptive systems (described according to a number of dimensions, such as physical functioning, pain and mental health) and valuations of those states. These include the Health Utilities Index (HUI) derived from SG (Horsman et al., 2003), the EQ-5D derived from TTO responses (Dolan, 1997) and the Sf-6d derived from SG responses (Brazier et al., 2002).

This methodology could be extended to preferences towards QoL states, rather than just health-related quality of life (HRQoL). Indeed, an aspect of QALY measurement that is often overlooked is that giving up years of life within the valuation process implies that everything of value to an individual has already been incorporated within the states being valued. This can be seen clearly with an example. In a TTO exercise if the individual is indifferent between state z for 10 years and full health for 20, each year in this state is valued at 0.5 QALYs. The change in health from state z to full health is being valued in terms of everything which is important about being alive, as it is not just the 10 years of health which is traded off, but 10 years of life. The use of the term HRQoL implies that other prudential concerns may be important in addition to QALYs. However, the valuation method implies a change in QALYs is valued against everything that matters to the individual about being alive. Consequently, since this method already captures the trade-off between quality of life and length of life it is ideally suited to extending beyond HRQoL into a measure of full QoL or well-being.

9 More commonly the VAS may be anchored between best and worst health state, then recalibrated to dead (0) and best health state (1). Because VAS does not involve any trade-off it can be argued not to present true preferences, consequently it is not recommended for use in cost-utility analysis by NICE.
10 For example, risk neutrality for SG, no discounting in TTO, and time neutrality.
11 The latter two are most widely used in the UK because representative public valuations are available for each health state for both of these measures. These valuations are currently being used by NICE to calculate the number of QALYs generated by new health care treatments that are used to inform recommendations about whether new technologies should be reimbursed by the NHS.
Summary of measuring well-being within preference satisfaction accounts

Measuring preference satisfaction adopts one of two approaches; comparing complete lives or aggregating individual preferences over moments of time. Comparing lives is similar to health-related QALY measures which also offer a weighting for current well-being and aggregation to a specified duration of time and different lengths of life. The aggregation of preferences can be proxied by income or consumption measures. These can be used for current well-being, well-being over a specified duration of time and lifetime well-being. Surveys on perceptions of desires satisfied, or the gap between expectations and attainment do not lend themselves to these clear time scale distinctions.

Measuring the summation of preference satisfaction requires adopting a stance on the contribution of unmet preferences for the measure of well-being. Measures may tap into quite different concepts depending on their treatment of unmet preferences. That a preference remains unmet is clearly important to well-being, otherwise there is no reason for supposing meeting a preference would matter to well-being. Where preference satisfaction is measured by income or consumption, preference formation is assumed to be exogenous, and can be thought of as a source of measurement error. At the other extreme within gap theories the focus is upon the unmet preferences as the important source of information for measuring well-being.

Current income only provides economic well-being, or that aspect of well-being which is attainable via the market. However, if it is assumed that economic well-being is, at all levels, positively correlated with total well-being this restriction should have little impact. This assumption has been dominant in welfare economics since Pigou (1932) who, whilst acknowledging the potential divergence between economic welfare and total welfare, argued that the “quantitative conclusions about the effect of an economic cause upon economic welfare will hold good also of the effect on total welfare” (Pigou, 1932, chapter 1, paragraph 10).

If this assumption is questioned, and it is instead assumed that the individual may have many important desires which cannot be met through the use of individual resources, and are not correlated with individual access to resources, then measuring complete satisfaction of desires requires the use of more than access to resources. This will be discussed in more detail in Chapter Five.

2.3 Mental state accounts

Mental state accounts are categorised here into hedonic, evaluative and subjective well-being (SWB) accounts. Feelings held by the individual may be important in all versions, but only within hedonism are feelings the only attribute of importance. Evaluations an individual makes
of her life are important to both the evaluative and SWB accounts. However, only within evaluative accounts are evaluations the only attribute of importance.

2.3.1 Hedonic accounts

2.3.1.1 Theories of hedonism

Hedonism is a mental state account in which well-being depends entirely upon feelings held by the individual. Hedonists traditionally take the view that pleasure and pain are the only things that are good or bad for anyone, and what makes these things good and bad respectively is their ‘pleasurableness’ and painfulness (Crisp, 2006). Well-being is therefore conceived as the total of all hedonic moments or the product of the intensity of pleasure (absence of pain) and its duration.

Experiences of pleasure and pain are not homogeneous, some being simple (such as physical sensations) and some being more complex (such as feelings of friendship or grief). However, most hedonic accounts require that pleasure and pain share a certain common property or quality which enables them to be compared.

Sensation hedonism compares feelings based on some shared mental state or sensation that is present in all pleasure and pain. This form of hedonism has faced strong objections, critics pointing to the apparent lack of a shared mental state amongst pleasurable (painful) experiences (Griffin, 1986; Sumner, 1996). Theorists therefore turned to alternative explanations for the shared quality of feelings.

An individual’s attitude towards a particular feeling may be assessed via their desires towards experiencing those feelings. This is the approach adopted by preference hedonism (the prime example of attitudinal hedonism). Desirable and undesirable feelings can, under preference hedonism, be aggregated to form a total picture of how well that individual’s life is going. Thomas Scanlon explains that within this account “the experience of living a life is made better by the presence in it of those mental states, whatever they may be, which the person living the life wants to have, and is made worse by containing those states which that person would prefer to avoid” (Scanlon, 1993: 186).

Fred Feldman endorses a type of hedonism where the degree of enjoying is separated from the intensity of the feeling. In order to overcome the objection that some pro-attitudes (e.g. desires) may exist towards mental states that are self-destructive or based on false premises, Feldman only considers a pro-attitude as valuable if it is directed at something that is pleasure-worthy. His ‘Desert Adjusted Intrinsic Attitudinal Hedonism’ weights the pleasure according to the
degree to which it arises from an object or state of affairs deemed to be worthy of pleasure and based on true state of affairs (Feldman, 2002: 14).

An alternative response to the absence of a shared psychological feeling of pleasure common to all experiences is to introduce a dimension of ‘quality’ of pleasure, as done by John Stewart Mill. In *Utilitarianism*, Mill adopts a hedonist framework yet he also incorporates the idea of there being higher and lower pleasures based on some objective idea of human potential against which individual lives can be judged (the judging to be done by those who have experienced both types of pleasures). Mill argues that since humans are capable of appreciating the superiority of mental over physical pleasures it is necessary to distinguish between higher and lower pleasures in their contribution to well-being. Consequently, pleasure is seen to vary in qualitative (different kinds) and quantitative terms (Mill, 1998 [1863]).

2.3.1.2 Measuring hedonic well-being

Measuring momentary well-being from a hedonic perspective is the amount of pleasure currently experienced by an individual, as could be identified by Edgeworth’s fantasised ‘hedonimeter’.\(^\text{12}\)

The exact level of pleasure experienced in any moment may be proxied by the individual’s judgement of the amount of pleasure they are experiencing. In sensation hedonism, where pleasure is understood to be a sensation that arises in all feelings, this may be colloquially thought of as the amount of happiness felt in any moment.

If the individual is unable to give a summary judgement a proxy for the extent of ‘pleasure’ experienced may be gained from knowledge of all positive and negative affects experienced. These may include a wide range of feelings such as, positive feelings like joy, a sense of connection, absorption, creativeness, excitement, sense of achievement, serenity, meaningfulness, and negative feelings like frustration, stress, sorrow, and grief. Some of these affects have an evaluative nature, such as achievement, however, to remain compatible with hedonism, they would refer to the aspect of the state that is a felt experience in the moment. Although some states, such as absorption, may not be traditionally seen as feelings it is still reasonable to ask the extent of pleasure involved in experiencing such states.

\(^{12}\) A machine capable of “continually registering the height of pleasure experienced by an individual” (Edgeworth, 1981: 101). Developments of physiological measures and brain imaging keep alive the possibility that a ‘hedonimeter’ may still be possible some time in the future. However, whilst these may link to certain types of feelings and emotional states it is questionable whether they can ever reflect the meta state of ‘pleasurableness’. 
To generate a single level of hedonic experience, an assumption is necessary about the extent to which the overall experience is pleasurable based on the combinations of types and the intensities of positive and negative affect an individual experiences. However, Diener has argued that measuring intensity is not necessary because the frequency and duration of positive and negative affect is a sufficient predictor of positive emotion (Diener, 1994, 2000).

To generate an overall level of the ‘pleasurableness’ of the moment it is necessary to aggregate from the presence of a range of positive and negative affects. Sensation hedonism assumes that pleasure and pain are located on the same spectrum, if not, an additional level of judgement is required in order to combine pleasure and pain to generate a single measure of hedonic experience.

From a preference hedonist perspective what matters is the strength of desire for the feeling which currently being experienced. Preference hedonism cannot therefore be fully measured without knowing an individual’s preference towards each feeling, such as the perceived desirability of feeling anger, tranquillity or nostalgia.

Within most hedonic accounts, an individual’s well-being over a specified duration of time is the linear aggregation of their momentary well-being, where each moment in time is treated equally.

The aggregation of momentary well-being to represent well-being over a specified duration of time is demonstrated by Kahneman et al. (1997). As noted above, in order for momentary well-being to be additive it must be possible to treat each moment separately, hence momentary well-being must be a “sufficient statistic” in that “all the information needed to evaluate the goodness of an episode must be incorporated in its utility profile.” (Kahneman et al., 1997: 390). Hence the order of moments must not be important. Also the duration and intensity of experience must be known, judgements of intensity must be comparable across individuals and a neutral position of experiencing neither positive nor negative affect must exist (Kahneman, 1999).

However, that some moments lived should be given more weight seems a common intuitive understanding. This may still be compatible within the equally weighting framework if the reason for that intuitive understanding is due to the time scale from which the moment is being assessed. Those moments which people intuitively may wish to give more weight to are likely to be those moments which impact on future experience. They may do so either because they will be remembered during future moments or contribute to the experience of future moments. Consequently, important moments could have a greater impact on total well-being than those which do not have subsequent consequences through influencing future affect.
Failure of remembered experience to represent an aggregation of momentary experience has been found in a range of experiments conducted by Kahneman and colleagues. Rather than aggregation of momentary experience experiments have found that duration of experience is often neglected and retrospective evaluations are best predicted by a Peak-End rule or the average of the most intense impression and the final impression (Fredrickson and Kahneman, 1993; Kahneman et al., 1993; Redelmeier and Kahneman, 1996). Ariely and Zauberman (2003) list a wide range of domains in which improving sequences have been shown to be preferred to numerically equivalent constant or declining sequences, including pain, discomfort, emotional episodes and life experiences.

However, to some extent it may be possible to reconcile the operation of a Peak-End rule with the validity of equally weighting moments. The retrospective assessment of a period of time may give additional focus to the peak and end since these are likely to have greater future impact upon our memories and future experiences. Of course it may simply be that our memory is at fault. If we use simplifying heuristics these may in many situations reflect the aggregate of momentary experience, but not in all. If this is the case, faulty recall will result in a source of measurement error.

The unweighted aggregate of moments shares strong similarities with the QALY approach, discussed above. In order to compare hedonic profiles representing different lengths of life the neutral position would need to imply indifference between living the moment and not living the moment from the perspective of the individual. The addition of a moment in time in which the individual’s experience falls below this neutral point would result in an overall affect which is lower than had the moment not been lived, from the perspective of the individual’s prudential value.

However, individuals may have different preferences towards experiencing states in the short term to experiencing them in the longer term. For example, in the longer term experiencing painful states may enable the attainment of other prudential values, such as the attainment of goals. Consequently, the simple aggregation and treatment of negative affect as worse than dead only makes sense within a hedonist framework in which mental states are the only thing which contribute towards well-being, and moments spent in predominantly negative states are valued as worse than dead. There may of course be other reasons for wanting to live the moment, such as the contribution made towards others during that moment, or other non-prudential values.

Time neutral weighting of moments assumes that an individual’s lifetime well-being is the aggregate of pleasure experienced in equally weighted moments of time. However, it may be possible to adhere to pleasure being the only thing that matters to an individual’s well-being.
without adhering to the requirement to equal weight moments over time. If the order of affective experience or the narrative quality of their life contributed to their well-being this would mean, in a strict sense, less pleasure being preferred to more pleasure.

From a hedonist perspective it is the actual feelings that are measured rather than the experience itself (i.e. the feelings arising from eating a meal, rather than the act of eating a meal) hence the impact of other information (previous meals, meals of others) will already have influenced the creation of the feelings. The appropriateness of equally weighting moments and considering the order of feelings to be not relevant are normative issues, but are compatible within a framework of hedonism. Hedonism therefore is well suited to aggregating moments of time to derive a long-term well-being profile. However, we require not just the theory but also the measurement instruments to meet these criteria. This would require any scale of intensity of positive and negative affect to be used independently of past levels of the scale. If the anchors of positive and negative affect change as people experience more extreme positive or negative events in life, their use of the scales may change (as discussed in more detail in Chapter Five).

Some existing measures which tap into well-being as conceived by hedonism are the Affect Balance Scale (Bradburn, 1969), the Positive and Negative Affect Scale (Watson et al., 1988), the Experience Sampling Method (Csikszentmihalyi and Larson, 1987) and the Day Reconstruction Method (Kahneman et al., 2004b).

The Affect Balance Scale (ABS) was designed to measure psychological well-being, defined as feelings of happiness and is concerned with “the pleasurableness or unpleasurable character associated with the experience” (Bradburn, 1969: Chapter 4: 54). A total score of the difference between 5 positive and 5 negative affects is calculated.

Attempts to combine positive and negative affect into a single score, such as the ABS, imply that positive and negative feelings lie in a one-dimensional scale, which has a mid-point, neutral position, of neither positive nor negative. The notion that positive and negative affect exist across one spectrum has been widely criticised. Bradburn (1969) was the first to demonstrate that positive and negative emotions form separate factors and are not polar ends of the same spectrum. Others have also found that positive and negative affect show a surprisingly low correlation and are related to different factors (Larson, 1987; Clark and Watson, 1988; Watson, 1988; Watson et al., 1988; Huppert and Whittington, 2003).

However, other researchers do find a single factor solution within happiness scales (Stones and Kozma, 1985) and a strong, inverse correlation between positive and negative affect. Variation in the strength of correlation seems to depend upon the scale and method used. For example,
Kammann et al. (1984) found the correlation between positive and negative affect in ABS to be -0.25, but -0.70 for positive and negative affect in the Affectometer 1. Benin et al. (1988) found the correlation to be stronger when using unequal item loadings and Green et al. (1993) found correlation to be higher when controlling for measurement error. Diener and Emmons (1984) found that the relationship between positive and negative affect depended on time frame. Diener et al. (1985) found that the relationship varied due to the interaction between affect intensity and frequency. If intensity was controlled for, the correlation between positive and negative affect was again strong. Larsen et al. (2001) found the simultaneous existence of positive and negative affect is mostly limited to emotionally complex situations and not likely in usual, day-to-day behaviour.

If positive and negative affect cannot be located on a single spectrum this does not preclude there being another scale, say of ‘pleasurableness’ on which positive and negative affects can be located. However, such a scale would require a judgement over how important positive affect was compared to negative affect and the weight of their respective contributions towards overall ‘pleasurableness’.

The Positive and Negative Affect Scale (PANAS) was designed to improve upon the validity and reliability of the ABS and treat positive and negative affect as independent dimensions. It can be applied to a range of time frames from right now, to the past week, month, year, or life in general. The scale asks the extent to which respondents have experienced 10 positive and 10 negative affects, providing two summary scores, and not giving any judgement about the overall ‘pleasureableness’ of a moment in time.

The ABS refers to the last few weeks, however, taken sequentially these could, in theory, represent a profile of an individual’s experience over a period of time. Similarly, it may be possible to generate a profile of affect scores from the PANAS. If the aggregate of PANAS scores representing weeks, for example, does not match a score representing months, this raises concerns over whether individuals report their own experiences in a time neutral manner. Differences may arise due to deficiencies in memory or if the time scale applied influences the use of the intensity scales. For example, Winkielman et al. (1998) show that reports of anger frequency depend on the temporal reference period. When respondents are asked ‘How often have you been angry during the past week?’ they report both trivial and important episodes. Yet when asked ‘How often have you been angry in the past year?’ only serious episodes are reported.

The Experience Sampling Method (ESM) also reports on the intensity of positive and negative affect, and aims to provide an overall assessment of a period in time. ESM involves signalling
subjects at random times throughout the day when they are supposed to record how they are feeling (happiness, frustration, worry etc.).

The Day Reconstruction Method (DRM) has been used to approximate to the more expensive and time-consuming ESM, and avoid potentially non-random missing observations, which arise due to the invasive nature of ESM (Csikszentmihalyi and Hunter, 2003). The DRM asks people to write a diary of the main episodes of the previous day and recall the type and intensity of feelings experienced during each event (Kahneman et al., 2004b). Kahneman and Krueger (2006) provide evidence that the results from the DRM provide a good approximation for those from the ESM. The DRM and ESM in their current form tap into how people feel, however, it may be possible to broaden the DRM measure and incorporate states beyond positive and negative affect including measures of engagement such as feelings of interest, focus and meaning. Depending on our understanding of pleasure this may be a hedonic account or may extent to include concerns beyond pleasure and pain.

To generate a measure of ‘pleasurableness’ from the ESM or DRM, a summary of how positive the moment was to experience needs to be generated from the responses to types of feelings and their intensity. However, there is little theoretical basis to draw on to support this calculation and hence calculating how intensely pleasurable a moment is from intensities of different affects can adopt a variety of different methods. Kahneman and colleagues have used a range of methods to summarise the net impact of each moment in time. One possibility is to take the difference between the average positive feelings (or the most intense positive) and the average negative (or the most intense negative) (Kahneman et al., 2004b). Once an overall summary of moments experience is gained, moments may be aggregated to form an integral of the day’s well-being. Rather than present an overall assessment of each individual moment, the proportion of time in which the most intense negative affect outweighs the most intense positive may be generated, referred to by Kahneman and Kruger (2006) as a ‘U-index’.

In summary, some of the options available for measuring momentary well-being which would be compatible with hedonism are the ABS, the affect information from ESM and the DRM. There is no direct measure for a proxy for preference-hedonism. Options for lifetime well-being involve duration weighted ESM, and DRM summary outcome measures as long as an assumption is made about the location of non-existence on the aggregated scale.
2.3.2 Evaluative accounts

2.3.2.1 Evaluative theories of well-being

In this section I outline the evaluative theory of well-being, paying particular attention to whether the account can deal with momentary well-being and aggregation of momentary well-being.

Within an evaluative, or life satisfaction, account of well-being an individual’s life goes well for her if she assesses her life positively. The individual’s overall life assessment will draw upon their own understanding of what makes life go well. The individual may hold a theory of well-being which differs from the evaluative account, they may think pleasure is the only thing which matters to their well-being, or the extent of desire satisfaction, or whether they are flourishing. Consequently, their assessment would be based on only one of these things, giving zero weight to other attributes. This does not imply that the evaluative account will adhere to each individual’s personal account of well-being, since it will remain entirely subjective. If the individual believes that the actual state of the world impacts on their well-being, rather than their perception of the state of the world, this will not be incorporated.

An individual’s assessment may include positive and negative feelings, whether they are happy with their lives more generally, or whether they feel they are meeting their idea of human potential. The terms ‘evaluative account’ or ‘life satisfaction account’ are not well used within philosophy, although a restricted version is supported by Sumner (1996).

Within Sumner’s evaluative account an individual’s life goes well for her if she has a positive attitude, in terms of affirmation or endorsement, towards her life, which in its fullest form has both a cognitive and an affective component, although the latter is still evaluative (Sumner, 1996).

Sumner describes the cognitive aspect as a positive evaluation of the conditions of your life, “a judgement that, at least on balance, it measures up favourably against your standards or expectations”. The affective component is a “sense of well-being, finding your life enriching or rewarding, feeling satisfied or fulfilled by it” (Sumner, 1996: 146), which is less cognitively demanding than a judgement about life as a whole. Sumner does not, however, provide details of exactly how measurement of well-being within this account would occur. It is useful, therefore, to explore the means by which this theory could give a measure of individual momentary well-being, well-being over a specified duration in time and lifetime well-being.
The affective component of the evaluative account is not a measure of the existence of feelings of pleasure in the sensation hedonist sense, but feelings of contentment and satisfaction. Positive and negative affect will contribute to both the cognitive assessment of whether life is going well, and to the generation of feelings of fulfilment. In both cases the contribution will be mediated by attitudes towards different types of feelings (suggesting greater similarity with attitudinal hedonism). Thrill seekers, for example, will place a high value on moments of excitement, while others may give more weight to moments of tranquillity. Some types of affect (such as anger or melancholy) may be positively assessed by some and negatively assessed by others.

People will also vary in the weight given to positive types of affect versus negative types of affect and how important they perceive it is to feel affect intensely. In addition, some negative states may be tempered if the individual perceives that it is an appropriate response to a set of circumstances, or a source of learning or a natural part of human existence. An assessment of feelings will also incorporate how the situation which gave rise to those feelings fits into broader life goals. For example, pain arising from a trip to the dentist is likely to be assessed differently to pain arising from falling over. Similarly, positive affect may be assessed more critically if it is perceived to undermine future plans or future positive affect.

The cognitive component is an overview of one’s life circumstances. A cognitive aspect, which could arguably be given greater emphasis than is done by Sumner, is an assessment of oneself as a person. This can be thought of as a cognitive assessment of whether one can personally endorse one’s values, ethics, beliefs, priorities, and whether one’s actions are consistent with one’s values. This assessment may also reveal itself in affective ways, such as feelings of self-esteem and contentedness.

Well-being in this context is the extent of positive affirmation, which is reflected both in our cognitive assessments of our lives and the person we are, and our feelings of satisfaction and contentment.

An individual’s evaluation of their life will be influenced by how closely their current life meets their expectations. At the extreme, being satisfied with life implies that there is nothing in your life that you would want to be otherwise and no unfulfilled wants (except those which are part of experiencing a desire to have something to aim for and look forward to). When seeking to assess my degree of satisfaction with my life I may consider my achievements and goals, whether I have challenges and goals to work towards and something to look forward to, whether I have future security, whether I have a role and purpose in life, whether I respect and feel connected to others, whether I both give and receive love from others, whether I have deep and
meaningful relationships, whether I have ongoing personal and spiritual growth, whether I approve of the person I am, and whether I have sufficient positive affective experience and acceptable levels of negative affective experience. As such it represents a far broader perspective on the nature of well-being than hedonism.

Different people may give different weights to these various aspects, for example for some people appraising their life as having achievement may be more important than giving love to others. An overall evaluative assessment allows individuals to give weight to the different aspects in line with their own priorities and judgements.

This appears to imply that they only have a well-being value when they are evaluating themselves, yet individuals do not continually assess their life. However, at times when individuals are not evaluating their life, the underlying components of their evaluation (e.g. their preferences, attitudes, feelings) exist and therefore a hypothetical evaluation of that person’s life may be made. In other words, an individual’s well-being at a point in time is determined by factors they would use if they were to evaluate themselves at that time.

Individuals are capable of making these kinds of general assessments and find it relatively easy to respond to evaluative questions about their life such as answering generic life satisfaction questions.

It may be objected that an individual may be deceived, or brainwashed, when making an evaluation. For Sumner, an individual’s assessment of happiness must be ‘authentic’ for it to count as well-being, where authenticity is being informed and autonomous (i.e. the assessment does not arise from manipulated desires or values). It is assumed that happiness is well-being unless it can be shown that an individual’s assessment is misinformed or non-autonomous. A normative judgement is required, therefore, about what constitutes being informed and autonomous.

Whilst the acts of evaluation and reflection generate momentary feelings (satisfaction with the situation, disappointment, approval etc.) it is the judgement itself that is the basis for well-being (although such feelings themselves may be included in evaluations).

For reasons mentioned above, having a momentary measure of well-being that can be aggregated is desirable. Is such a measure available to evaluative accounts? Within hedonism, momentary well-being is taken to be units of pleasure, therefore aggregation is possible across any time period. However, within an evaluative account the basic currency of well-being is an individual’s evaluation of their life. There is no obvious arithmetic way of decomposing such an evaluating judgement into units of momentary well-being. The corollary to the amount of
pleasure experienced within a moment would be ‘in my evaluation, this moment contributes x to my overall well-being’.

However, this type of decomposition may lead to inconsistency. For example, I may judge that one particular second (t₁) makes no contribution to my well-being. Similarly I may judge that seconds t₂–t₁₀₀₀ may make no contribution to my well-being, when considered individually. Hence t₁ to t₁₀₀₀ appears to make zero contribution. Yet, this very same period of time, when evaluated as one interval spanning 1000 seconds may make a positive (negative) contribution. Consequently, evaluations are hard to attach to any specific moment in time.

Furthermore, if an assessment of a single moment incorporates a prediction of future positive experience (beyond the impact that prediction actually has on the experience in the moment), momentary well-being will be inappropriate as it is not an outcome measure but rather a prediction of experience for an unspecified period of time.

If momentary well-being could be identified then lifetime well-being could be treated as an aggregation of momentary well-being. This could adopt a similar form to the aggregation process within hedonism, where each individual’s lifetime well-being was the unweighted sum of their well-being associated with each moment, and the aggregate was independent to the order of those moments. To compare lifetime profiles with different lengths of life a scale of evaluation would need to encompass evaluation which is equivalent to not being alive.

Incorporating a non-existence state within hedonism raised concerns that whilst the individual may value particular moments as worse than dead, they may change this evaluation if this state continued into the longer term and such things as achieving goals and maintaining relationships were important beyond affect states. Within an evaluative framework, evaluation of life overall should incorporate everything that is important.¹³

However, if momentary well-being cannot clearly be identified, then aggregation of moments will not be possible. Rather some proxy measure which attributes the same level of well-being to each moment as the individual assigns to the longer period of time may be necessary. In aggregating moments, as with hedonism, the assumption of equally weighting moments has an advantage in that each moment is lived and experienced for the same length of time and no one time point, or vantage point, is granted authority over any other time point. It is, as argued by Broome, the most reasonable default position to adopt and seems the best position to adopt from the perspective of measuring well-being for public policy purposes.

¹³ Non-prudential values, which may or may not be included within life evaluations (see Chapter Five), could be a source of deviation.
Nevertheless, individuals themselves may not consider well-being to be an aggregate of moments in time. If an individual adopts a conception of well-being in which the narrative or story of their life is dominant, it may be possible for them to experience low levels of momentary well-being through most of their life, then high levels of well-being prior to death and evaluate their lifetime well-being as high. Similarly, if all an individual’s goals are achieved at the very end of life, and an individual perceives well-being as fulfilment of goals, lifetime well-being may be considered by the individual to be high. A theory which aggregates momentary well-being will need to address why preferences and subjective assessments of momentary well-being are taken as important but not preferences and subjective assessments of lifetime well-being. There is some inconsistency with the aggregation of momentary well-being as a measure of lifetime well-being from the perspective of an evaluative account if it does not marry with an individual’s own longer term assessment.

Equally, there appears no reason from a policy perspective to privilege one moment of time (say an end of life assessment) over any other. For example, if an individual spent 60 years of his life with a low self-endorsement of that life, but in his final year of life he gave a positive endorsement, he would seem to have lower lifetime well-being than an individual who gave an equally positive endorsement in his final year of life, but had high self-endorsement over the previous 60 years.

In summary, the evaluative account is not theoretically suited to identifying momentary well-being (unlike hedonism). Momentary well-being can only be for a period of time that is fixed in an arbitrary fashion. Consequently, momentary well-being is a purely artificial derivative of well-being over larger periods of time, hence it should be treated with caution, and be subject to further research.

The assumption that moments can be equally weighted is not fully compatible with a purely subjective evaluative account. However, an evaluationist inspired account that adopts this framework would seem a reasonable simplification for policy purposes.

Furthermore, the means by which an anchor point can be created to enable comparisons of different length profiles remain unexplored. These issues will be discussed further in Chapter Nine.

### 2.3.2.2 Measuring evaluative accounts

There are a variety of options for both a direct or indirect measure of an evaluative assessment of an individual’s life. However, it is often not clear whether these are picking up a momentary assessment of well-being or an assessment over a longer duration.
Measures compatible with an evaluative account are far less specific over the time scale to which they apply, both in the language of the questions (some do not even give a specific time scale) and in the likely time scale which subjects would use when reporting about the general state of their life (even when a time scale is specified within the question there is little evidence to suggest that this is the time scale used by respondents). Lifetime well-being from an evaluative perspective could occur at the end of life. However, this is of little benefit to policy decisions, unless it could be accurately predicted from current measures.

Respondents own assessment of their life or certain aspects of their life can be collected in a variety of ways. Firstly directly, using individual surveys and interviews. Such surveys usually present respondents with evaluative style questions about their life, and a choice of responses using some type of Likert scale (Likert, 1932), with either numbers, indicative words, phrases or sentences, rungs of a ladder or even smiley faces (Andrews and Withey, 1976).

Subjects may be asked single, evaluative style questions which can focus on overall happiness or life satisfaction. An alternative is to consider the perceived difference between the individual’s actual life and their ideal life.

Life satisfaction may also be assessed via multiple questions, such as the Satisfaction with Life Scale (SWLS, Diener et al., 1985). This asks levels of agreement with five statements relating to how satisfied they are with their life and how closely it has lived up to their expectations. Questions include “If I could live my life over, I would change almost nothing” and “My life is close to the ideal”. A development from the SWLS is the Temporal Satisfaction With Life Scale (TSWLS, Pavot et al., 1998) which uses the same structure as the SWLS but reports past, present and future separately.

Rather than question people about their lives in a holistic sense, an evaluation may be built up from an individual’s assessment of different domains of their life. “Satisfaction with life as a whole can be seen as an aggregate concept, which can be unfolded into its domain components” (van Praag et al., 2003: 3).

---

14 For example, overall happiness (‘Taking all things together, would you say you are…very, quite, not very, not at all happy’ (World Values Survey)), or life satisfaction (e.g. ‘On the whole, are you …satisfied with the life you lead?’ 4 very satisfied, 3 fairly satisfied, 2 not very satisfied, 1 not at all satisfied, Don’t know’ (Eurobarometer), or ‘how things are’ (e.g. ‘Next are some questions about how you see yourself and your life. Taking things all together, how would you say things are these days?’ where 1 denotes “very unhappy” and 7 denotes “very happy” (NSFH, USA) or similarly ‘How do you feel about your life as a whole,’ with seven response options anchored at delighted to terrible (Andrews and Withey, 1976)), or contentment (e.g. ‘How contented have you been on your life? Do you feel that everything is as it should be and there is no need for change? Or do you feel dissatisfied and discontented?” where degrees of contentment are described by verbal labels (Ojanen’s Descriptive Visual Analogue Scale)).
This requires, firstly knowing which domains of life are important, and secondly knowing how success in one domain contributes towards the overall evaluation. Externally determining which domains to consider and their contribution makes this measure no longer fully subjective. Instead it becomes a measure of how individuals feel about aspects of their lives that have been taken (externally) to be important. Here once again, we can see the evaluative account taking on normatively imposed characteristics.

Domains are usually established by: (i) simply drawing on the literature; (ii) responses to questions asking people what matters in their life; or (iii) beginning with a broad spectrum and narrowing down by factor analysis. Schalock (2004) finds that eight life dimensions are prominent within the literature: emotional well-being, interpersonal relationships, material well-being, personal development, physical well-being, self-determination, social inclusion, and rights. Domains used within the literature show considerable variation. Hagerty et al. (2001) note that domains have included a wide range from natural endowment, sexual functioning, personal growth, absence of depression, use of public services, and meaning of life.

Researchers have tended to assume that overall well-being can be represented by a linear additive combination of domain satisfactions (Headey et al., 1985; Cummins, 1996). However, this faces two concerns, firstly whether this uses an appropriate assumption of causality and secondly, whether the linear additive model is appropriate.

Overall evaluation may not be an aggregate of individual domains in a ‘bottom up’ manner, but rather generates a ‘top-down’ influence on domain satisfactions. The strength of the relationship between different domain satisfactions has been seen as support for some ‘top down’ influence (Diener, 1984), others pointing to the impact of stable personal characteristics on domain satisfactions (Costa and McCrae, 1980). Controversy remains (see Scherpenzeel and Saris, 1996), although it appears that the relationship between domain satisfaction and overall satisfaction operates in both directions, is in part related to stable personality traits, and the relative strength of ‘top down’ effects varies from domain to domain (Headey et al., 1991). The potential disagreement across which domains matter amongst different researchers suggests a selection of chosen domains may not fully represent what all individuals judge to be important in their lives.

The linear additive model assumes that domain satisfactions are commensurable. However, aggregation across domain satisfactions will be inappropriate if they are non-commensurable components of a multi-dimensional concept of well-being (Rojas, 2006).
An example of an instrument based on this domain satisfaction approach is the *Australian Unity Personal Well-Being Index* (PWI, Cummins et al., 2003) which originally recorded respondents satisfaction with the seven domains; standard of living, health, achievements in life, personal relationships, how safe you feel, community connectedness and future security. These domains were chosen such that each domain can be thought of as an unequivocal indicator (i.e. non-instrumental), and were found from a review of 27 QoL definitions and findings from large population studies asking which domains are important to people (Cummins, 1997).

The domains are “theoretically embedded, as representing the first level deconstruction of the global question: ‘How satisfied are you with your life as a whole?’” (The International Well-being Group, 2005: 8). An additional theoretical constraint was applied that domains should be amenable to both subjective and objective measurement. Focusing on the sub-components of overall assessment provides additional, potentially useful information, which can highlight particular domains of interest for policy. The PWI weights each domain equally, taking the average of the domain satisfactions. In 2006 an eighth domain, satisfaction with religion or spirituality, was added, reflecting the nature of the PWI as work in progress (Cummins, 2007).

An alternative to the use of equal weights across the domains is to use weights determined from statistical analysis that links overall life satisfaction to domain satisfactions.15 The average weights from the population or subgroup used would then be applied at the individual level. However different domains may be valued differently by different people, or at different stages of the life course.

An earlier variant of the PWI, the Quality of Life Inventory (QOLI), aimed to take account of this variation (Cummins, 1997). The QOLI assesses an individual's QoL through self-report of the importance they attach to each of 16 domains as well as their satisfaction with each domain. For each domain, importance scores are multiplied by satisfaction scores and these scores are summed to determine QoL for each individual. This way the QOLI attempts to incorporate a subjective component of how important each domain is to an individual. The Quality of Life Profile developed by the University of Toronto also uses this approach.16

However, the PWI moved away from this method since it requires the assumption that moving one point on the scale of importance is equivalent to moving one point on the satisfaction scale (or any fixed ratio between scales), and there is no theoretical or empirical basis for such an

15 A similar approach is used by the Economist to generate their Quality of Life index (The Economist, 2005) but rather than using domain satisfactions they use other objective indicators and regress these on life satisfaction responses to generate weights for the importance of the factors, which they apply to all countries.

assumption (Trauer and Mackinnon, 2001). Importance weights are seen as difficult to interpret with undesirable psychometric properties and unnecessary because extreme scores on the domain satisfactions are associated with higher importance weightings suggesting the satisfaction scores incorporate judgements of importance (Trauer and Mackinnon, 2001; Wu and Yao, 2006; Hsieh, 2004).

Evaluative accounts can be measured through survey questions. Where these are single questions the individual weighs up all that is important in assessing their life. However, multiple questions, such as domain satisfactions, must find some way of weighting different aspects of assessment such that they closely represent the individual’s hypothetical overall evaluation, or adopt a normative stance on the importance of the different aspects.

### 2.3.3 Subjective well-being

#### 2.3.3.1 Theories of subjective well-being

Psychologists have shown increasing interest in measuring what they describe as subjective well-being (SWB). This is taken to be an umbrella term for how people think and feel about their lives, usually presented as comprising of three elements “people's longer-term levels of pleasant affect, lack of unpleasant affect, and life satisfaction.” (Diener, 1994: 103). SWB looks very similar to an evaluative account of well-being, so it is useful to distinguish the two at a conceptual level.

Those who use the term SWB do not always treat this concept as an account of how well an individual’s life is going, but as an attribute of importance, and element in a substantive good account, which may contain other elements such as objective indicators. The distinction between measuring a part of well-being or complete well-being will be important in interpreting empirical findings. For example, if it is shown that increased income does not increase SWB, if SWB is equated with well-being, this is very important for policy makers. However, if SWB is only part of well-being, knowing that the SWB component is not related to income is interesting, but may have limited policy implications. SWB can refer to a family of scales which are about an individual’s experience and their life, but do not necessarily attempt to present a holistic picture of how well that life is going. As can be seen in the definitions below, SWB is occasionally distinct from well-being and occasionally synonymous with well-being.

---

17 An additional requirement that feelings are “justified” and based on reality may also be added (Diener et al, 2003: 207), suggesting similarities with Sumner’s authentic happiness.
Relatedly, whilst an evaluative account would conceive of well-being as an overall assessment of life into which positive and negative affect are important contributors, SWB can be presented as containing separate components, evaluation of life, positive affect and negative affect. The difference between these two is subtle, especially at the level of measurement. We may hold a fully evaluative view of well-being, yet think that measurement of well-being can be better understood through supplementing with measures of positive and negative affect, where an individual’s evaluation of their life is reflected both in their stated evaluation response and their experiences of positive and negative affect.

**Current formulations of SWB**

The approach to determining what comprises well-being typically applied by physiologists is to draw upon variability within data which broadly relates to QoL. For example, Athaud-Day et al. (2005) tested the construct validity of SWB by looking at responses to a range of evaluative and affect-based scales. They found a three factor solution which is seen as supporting the ‘tripartite’ theory of SWB (Athaud-Day et al., 2005: 465).18

Where SWB is described as a cognitive evaluation of life plus positive and negative affect, it appears that feelings contribute to well-being above and beyond an individual’s assessment of the those feelings. As a combination of overall assessment (usually in the form of life satisfaction) and positive and negative affect this would suggest that SWB combines aspects of hedonism and an evaluative account, without a coherent structure to explain what makes positive affect and overall assessment good for an individual. This is a mental state account with three non-commensurable items.

However, life satisfaction and affect have been described as separate components of SWB which remain part of a higher order construct (Diener, 1994: 108), yet it still remains unclear whether this higher order construct is a global evaluation of life, or something separate to that.


“Subjective well-being is people’s subjective evaluations of their lives, and includes concepts such as life satisfaction, pleasant emotions, feelings of fulfillment, satisfaction with domains such as marriage and work, and low levels of unpleasant emotions” (Diener and Scollon, 2003: 4).

---

18 A potential problem of relying upon this approach without a clear theoretical foundation of what should be included is that it replies upon the initial data and factor analysis may identify a separate factor (such as sexual functioning), which is an input into other domains (such as life satisfaction).
However, Diener and Seligman (2004) present well-being (rather than SWB) as four slightly different components.

“Well-being, which we define as peoples’ positive evaluations of their lives, includes positive emotion, engagement, satisfaction, and meaning” (Diener and Seligman, 2004: 1).

In Diener, Scollon and Lucas (2003: 192) SWB is presented as a conceptual hierarchy reflecting a general evaluation of life, in which global life judgements are equated with life satisfaction and contribute towards SWB along with pleasant emotions, unpleasant emotions and domain satisfactions. How global life judgements differ from general evaluation of life remains unclear.

The term SWB is used within the literature to refer to subjective measures that relate to or pick up some aspect of how well an individual’s life is going. It is not reserved for measures which present a complete picture of individual well-being. Keyes describes SWB as “multifactoral, multidimensional concept” and notes that researchers have used 13 dimensions within the umbrella of SWB, including positive affect, avowed life satisfaction, avowed happiness, six dimensions of psychological well-being and five dimensions of social well-being (Keyes, 2006: 4). The wide range of SWB constructs corresponds with a wide range of measurement instruments.

### 2.3.3.2 Measuring subjective well-being

Various instruments aim to measure SWB as a combination of important subjectively held thoughts and feelings. All the measures discussed under evaluative accounts could also be categorised as measuring SWB. The three scales briefly discussed below are examples of scales which aim to measure SWB but can be distinguished from life satisfaction and general evaluations on the basis that they do not rely on the presence of a single, higher order construct of overall evaluation.

The First European Quality of Life Survey considers SWB as the “main provider of individual well-being” (Böhnke, 2005: 95). It includes a single question on overall happiness, one on overall life satisfaction and additional questions on sense of belonging and feelings of alienation. The latter two being seen as dimensions of SWB separate from the cognitive (life satisfaction) or emotional (happiness) aspects capturing “a general attitude to life, encompassing more than private matters and referring to surrounding societal circumstances as well” (Böhnke, 2005: 16). In addition, satisfaction with a range of domains is also included, although in this case these are seen as contributors to SWB rather than components of SWB. A single indicator of perceived quality of life is constructed in several ways through combinations of response to the happiness, life satisfaction and social alienation questions.
The Global QOL Assessment incorporates the SWLS, but uses an additional 11 questions to assess feelings and thoughts the subject holds about their life, their degree of satisfaction, how closely it matches their aims and goals, and positive and negative feelings held.

The Life Satisfaction Index (LSI) (Neugarten et al., 1961) measures the individual's evaluation of his or her psychological well-being. Life satisfaction is defined as including five components: zest (versus apathy), resolution and fortitude, congruence between desired and achieved goals, positive self-concept, and positive mood tone.

In addition to combining measures that tap into evaluative and hedonic well-being, SWB can be measured indirectly. This method overcomes problems which arise from self-report, such as incentives individuals may have to present themselves in a particular light. From considering certain behaviours, appearances and states, it may be possible to infer something about how the individual experiences or thinks about their life, although it will be unclear whether they reflect one aspect of SWB, such as positive affect, or an overall evaluation of life.

**Indirect measures**

Behaviours such as suicide attempts, heavy alcohol or other drug intake, or migration could be taken to infer that an individual is unhappy with at least some aspect of their life.\(^{19}\) However, multiple interpretations could be made from such behaviours.

Simulated behaviour has also been used as an indirect assessment of how well life is going for an individual. For example, studies have looked at peoples’ ability to recall as many good life events and then as many bad events as possible in a short time period (Seidlitz and Diener, 1993), or word tasks which explore the speed with which people recognise positive words and how they complete word stems (Rusting, 1998).

Another means of avoiding self-report is to use assessments from informants, such as family, friends, or external observers. This may be done by coding facial expressions (Ekman and Friesen, 1977), looking at the frequency of smiling, and types of smiles, specific muscle movements that react to emotional stimuli or a more general assessment of how happy a person appears to be (Diener, 1994).

Indirect assessment could also be taken from physiological correlates and biomarkers. Emotional responses to situations have been measured by the presence of stress hormone, cardiac variability, heart rate, blood pressure, bodily temperature, finger temperature, skin

\(^{19}\) In this case it is assumed that evaluation rather than preference is revealed through behaviour and choice.
conductance (Cacioppo, 2004; Clow, 2004) and brain activity recorded by electroencephalograms, PET scans and functional MRIs (Davidson, 2004). These types of measures are likely to pick up the presence of moods, and as such are potentially a better measure of well-being from a hedonic perspective than a complete measure of SWB. However, research is still in the early stages and it is not clear exactly what these biomarkers are picking up.

The presence of serious mental illness or psychotic disorder is another sign that someone is not happy with their life. Similarly, mild psychopathology, particularly depression and anxiety can be used to infer low levels of positive affect, and also negative life evaluations more generally. They may also be used to infer a low level of functioning, positive psychological well-being being seen as an essential component of the flourishing life. Consequently, mental health measures may proxy for well-being from hedonist conception, an evaluative conception, a multi-item SWB conception and a flourishing conception.

There are various measures to diagnose mild psychopathology. At the clinical level a distinction is made between depression and anxiety, each condition being associated with different functional impairments and disabilities, hence assessed separately (Spitzer et al., 2006). Within primary care self-report, brief diagnostic measures have been developed for depression (e.g. PHQ-9, Spitzer et al., 1999) and anxiety (e.g. GAD-7, Spitzer et al., 2006). However, at the population level the correlation between depression and anxiety often results in the use of a single scale to identify generic mental health problems.

Those which have been used within the SWB literature as proxies for well-being tend to be those which are found on large scale surveys. Commonly used measures of depression and anxiety include the General Health Questionnaire (GHQ) and the Centre for Epidemiological Studies Depression Scale (CES-D).

The GHQ was developed as a screening tool to measure mild somatic and psychological symptoms (Goldberg and Williams, 1998). The GHQ-12 has 12 questions that tap into both evaluation aspects (e.g. thinking of oneself as worthless) and current feelings, covering both positive and negative affect. The GHQ-12 is available in all waves of the BHPS and will be discussed in more detail in the next Chapter.

The CES-D was designed by Radloff (1977) as a 20-item measure of depressive symptoms in the general population. It measures current levels of depression focusing mainly on the affective component and includes positive as well as negative items, although also contains symptoms such as poor appetite making it less than ideal as a hedonic measure.
The extent of positive mental health may more closely reflect well-being for the majority than the presence or absence of negative mental health states as determined by measures which are designed as diagnostic tools. Two measures which aim to report the extent of positive mental health are the Affectometer 2 (Kammann and Flett, 1983) and the Warwick-Edinburgh Mental Well-being Scale (WEMWBS, NHS Health Scotland, 2006; Parkinson, 2006).

The Affectometer 2 is a 40-item measure developed to assess positive well-being. Ten underlying concepts were identified as uniquely contributing to mental well-being, and each is tested by 4 items: a positive sentence (e.g. “I have been thinking clearly and creatively”), a negative sentence (e.g. “My life seems stuck in a rut”), a positive adjective (e.g. “understood”) and a negative adjective (e.g. “lonely”). Responses to negative items are summed and subtracted from the sum of positive items, hence mental health status is perceived as the degree to which positive feelings and attributes outweigh negative feelings and attributes.

The WEMWBS developed from the Affectometer. It asks for time spent in 14 positive states over the last two weeks. This scale aims to cover most aspects of positive mental health including hedonic and eudaemonic perspectives: positive affect (feelings of optimism, cheerfulness, relaxation), satisfying interpersonal relationships and positive functioning (energy, clear thinking, self acceptance, personal development, mastery and autonomy).

In summary, measures drawn from an evaluative account or SWB version of well-being may be direct, survey questions, either single or multi-item with a range of language and question styles to pick up an individual’s generic assessment. Alternatively, they may be indirect, using actual or stimulated behaviour, observer reports, and physiological or psychological health states.

However, the distinction between positive appraisal of one’s life in a moment in time and summation of those assessments over time is not reflected in the measurement scales available. In terms of using SWB to represent momentary well-being, well-being over a specified duration of time and lifetime well-being, similar issues arise as those facing hedonic and evaluative accounts. The evaluative aspect creates problems for momentary well-being and the lack of an anchor to represent non-existence creates problems for comparing profiles with different lengths of life.

There is no measure which adequately reflects how valuable one moment in time is to live, incorporating an assessment of affect and life appraisal as experienced in that moment. Furthermore, whilst life satisfaction and global questions may specify a time scale from which the assessment should be made, it is not clear if the contribution of the past and the future

Weighs more heavily in responses than in their actual contribution to well-being during the specified period. In estimating well-being over a specified period of time it may be possible to make assumptions about the duration of the assessment, and present a profile of life satisfaction. For example, if life satisfaction is asked annually, with reference to the previous 12 months, it may be reasonable to attribute this level of life satisfaction to the previous year, thereby gaining a profile of life satisfaction across the time period of the survey.

2.4 Hybrid accounts

It may be possible to combine the different accounts of well-being into a multidimensional account, in which the different aspects, such as having one’s desires met, experiencing happiness and fulfilment, having knowledge and engaging in rational activity combine to make an individual’s well-being (Scanlon, 1993). Each aspect therefore represents only part of what it is to understand an individual’s well-being. For example, Keller (2004) sees goal attainment as important to an individual’s well-being but only one part of well-being.

However, combined multidimensional versions of well-being will not provide a shared quality between these attributes which explains their contribution towards individual well-being. Such hybrid accounts can be seen as substantive list accounts, in which attributes on the list are included because they are intuitively viewed as having non-instrumental value for the individual.

Substantive good lists that contain only subjective elements are identical to SWB accounts that incorporate more than either evaluation or an aggregate of positive and negative affect. Consequently, those measures listed under SWB accounts could equally be included within substantive good lists. Indeed, to do so would highlight the lack of formal theory behind many of the versions of measurement of SWB.

Other measures are part subjective, in the sense of the judgement being made by the individual, but not fully in the sense that some objective judgement is also included. Good examples of such measures are the World Health Organisation Quality of Life measures (WHOQOL) and the CASP-19 (Hyde et al., 2003).

The WHOQOL is in part compatible with a SWB account of well-being, and presents a combination of different objectively important attributes the levels of which are determined subjectively. The long version (WHOQOL-100) covers 24 dimensions of QoL that have a high level of international consensus, plus an overall evaluation of life. A shortened version
Chapter 2

Measuring well-being

( WHOQOL-BREF\(^{21}\) containing four dimensions, physical, psychological, social, and environmental, is also available (The WHOQOL Group, 1998). QoL, which appears interchangeable with well-being, is a broad ranging concept affected by the person's physical health (e.g. *to what extent do you feel that physical pain prevents you from what you need to do*), psychological state (e.g. *how well are you able to concentrate*), social relationships (e.g. *how satisfied are you with your personal relationships*) and their relationship to salient features of their environment (e.g. *how healthy is your physical environment? do you have enough money to meet your needs?*). As can be seen from the example questions, the WHO measure incorporates both outcomes (pain) and inputs (having enough money). This brings some ambiguity to the concept being measured. The measure also captures aspects of capability and opportunity (e.g. *to what extent do you have the opportunity for leisure activities?*) and functioning (e.g. *how often do you have negative feelings such as blue mood, despair, anxiety, depression?*).

The CASP-19 was developed to investigate the quality of life of older people in the UK, focusing on their common psychological and physical needs. The initials stand for four dimensions: Competence, Autonomy, Self-realisation and Pleasure. The attributes are seen as commensurable and an aggregate score is given. Although this scale originated from a needs perspective (see Appendix III) it focuses on subjective outcomes rather than inputs therefore may be best thought of as based on a hybrid, subjective approach. The CASP-19 has been collected within the BHPS data will be looked at in more detail in Chapter Three.

2.5 Summary and discussion

This chapter has attempted to show, within each account of well-being, how momentary and long-term measurement of well-being could be operationalised and offer some options for actual measurement of well-being over these different type periods. Some of the popular existing instruments (in the English language) used to measure well-being in health, social research and public policy which can be aligned to the different accounts of well-being have been considered. This by no means covers all the possible measurement instruments available to measure well-being, rather gives a flavour of the type of instrument which could be used if a particular theoretical stance is adopted.

Preference satisfaction as a conception of well-being has widespread support but from the perspective of measurement there are many unresolved theoretical issues. Although offering a

theory of what is of value for an individual, aggregation of preferences does not provide a clear structure for the measurement of lifetime well-being, nor momentary well-being.

Some unresolved issues include: Which preferences count? If only a sub-set of actual preferences count how are they to be distinguished in measurement? How are preferences to be aggregated? How is intensity of a preference to be judged - would length of time the preference is held be most influential or mental craving associated with the preference or sacrifice made in order to meet the preference or cognitive assessment of the importance of the preference for ones life? How do unmet preferences contribute to well-being? How important are preferences related to altruistic concerns? How should we deal with inconsistent preferences or those that change over time?

Measurement options within preference satisfaction centre upon the use of income and consumption. Both can be clearly measured across the range of time periods. However, the use of income and consumption rests on a range of assumptions: that preferences can be treated as exogenous hence the number of un-met preferences can be excluded from measurement; that income is allocated in a predictable manner across the household members; and that preferences which the individual has no opportunity to reveal are unimportant.

The theoretically superior measure of preferences towards alternative selves has not been operationalised although shares similarities to QALY measures which give preferences towards alternative HRQoL states. Both income and QALY measures can generate a well-being measure for a moment in time, duration of time and lifetime. However, in both cases, the glossing over of issues such as changing preferences over time, leads to measures that appear more coherent and theoretically grounded than closer inspection supports.

Hedonism lends itself to a single measure of ‘pleasurableness’ of experience, which can be conceived of at the momentary level, aggregated across a duration of time, and, as long as the scale captures the point which is equivalent to not-living, across different length life times. However, no discernable scale of ‘pleasureableness’ has been generated. It has been suggested that there are two independent dimensions of affect, positive and negative. If so, there is no clear conceptual basis for weighting these two dimensions. This then undermines the neatness by which momentary well-being can be generated. Without a clear summary of the intensity of pleasure experienced within a given moment, hedonic measures such as the DRM and ESM rely upon some debatable assumptions about how different affects experienced should be summarised. Furthermore, positive and negative affect summary measures may still not adequately measure ‘pleasurableness’ which may be a far broader concept if it incorporates individual preferences towards experiencing certain affects or intensity of affects.
Evaluative versions of well-being pass the task of determining well-being over to the individual, as such the weighting of affect or any other factors considered important is done by the individual. However, identifying momentary well-being is problematic as individuals do not isolate a particular moment of time to generate their assessment of their life. Consequently, a bottom up aggregation of moments is not conceptually compatible with this account. Evaluation is possible across larger units of time. Assessing different length lives requires the location of indifference between experience and non-experience within self-evaluations and where this falls is not yet clear.

Some measures, for example the CASP-19 and WHOQOL, do not neatly fit into any account hence are categorised here as hybrid. These measures are based on the premise of well-being being multi-dimensional, incorporating both objective and subjective aspects, and appear uncommitted to any formal theory of well-being. Whilst this has advantages of appearing to be inclusive and accommodate a range of ways in which an individual’s life can be thought to improve, it has the disadvantage of potential theoretical confusion. The use of a single index from such scales will either be *ad hoc*, or will require a higher construct of well-being.

Hence despite the theoretical clarity of most theories of well-being there is much less clarity surrounding how these conceptual frameworks translate into momentary well-being, well-being over a period of time and lifetime well-being, and much less still on how these concepts could support actual measurement of well-being across these different time periods.

This overview has revealed considerable variability in both what is held to be the attributes which constitute individual well-being, and the reasons why these attributes are taken to be individual well-being. This is unfortunate since if the concept of well-being is to drive public policy decisions it requires some stability and the stability of the concept used may be threatened unless there is widespread support and agreement upon what constitutes well-being. However, in the absence of consensus, clarity of the conception of well-being being used will be essential. Only then will there be a framework for understanding what should be measured, whether the well-being measure is valid and accurate, and what limitations the well-being measure may have. It is impossible to ask if a measure is a good measure of individual well-being without first knowing what is meant by individual well-being.

Although accounts of well-being offer different perspectives on what is of value, the different starting points may disguise the fact that certain versions of each account share very similar end points. For example, Tefler (1980) links the conception of having a favourable attitude towards one’s life and desire fulfilment by claiming that desire fulfilment is necessary to be able to positively evaluate one’s life. She also links individual flourishing (or *eudaemonia*) and
hedonism, arguing that the existence of *eudaemonia* is likely to indirectly maximise affective happiness (Tefler, 1980: 41). Pleasure may well be a by-product of flourishing, but empirical research is needed to establish if they are proportionately related.

Many accounts incorporate experiences of positive and negative affect as important. Within preference satisfaction this is because we are likely to desire states that bring positive and avoid negative affect. Within an evaluative account or objective list account this is because we are likely to hold consensus views that positive feelings are part of realising human potential. Furthermore, positive and negative affect may mirror well-being from a flourishing perspective, if, as Tefler claims, realising human potential is the source of maximum affective experience.

How much variation there may be in practice from an account of well-being including only feelings and accounts which incorporate other aspects depends on how much will not be picked up, or closely mirrored, by affective experience. In the case of an evaluative account this will be the extent to which individual assessments of life depend upon something other than feelings or the extent to which those things the individual counts as relevant to their assessment generate proportionate positive and negative affect. In the case of preference satisfaction accounts this will be the extent to which people have desires towards states of the world that are not captured by the feelings the states of the world generate. In the case of objective list accounts this will be the extent to which objective circumstances are reflected in feelings.

Preference hedonism shares similarities to a preference satisfaction account of well-being in that the contribution of an event to well-being depends upon an individual’s preferences towards that event. However, preference hedonism is restricted to preferences towards the feelings generated by the event, and preference satisfaction by a combination of feelings and cognitions associated with the event.

The qualitative hedonism of Mill appears closer to an idealised desire satisfaction account or an objective list account of well-being than it does to hedonism. If certain pleasures are considered of greater quality than others, the means of establishing the quality of a pleasure draws upon objective traits extending beyond the individual’s current experience. These quality judgements arise either from idealising preferences or from objective judgements that some pleasures have more intrinsic worth, or allow individuals’ experience to be closer to human flourishing.

Idealised preference satisfaction accounts are likely to give very similar components of well-being as substantive good accounts (Griffin, 1986). Indeed, versions of idealised desire accounts are likely to come much closer to substantive good list, and flourishing accounts than to actual desire satisfaction accounts. If actual desires are not accurately driven by maximising positive
and negative affect, or life satisfaction, or human flourishing, well-being according to that account will differ from all others.

The similarities suggest the potential for a list of attributes that would maximise individual well-being, which could be supported by different theories of the cause of well-being. This is pointed out by Sumner,

“Theories which start off at opposite ends of the subjective/objective continuum tend to be pulled towards the middle, and therefore towards one another, in the process of responding to hard cases and avoiding apparently decisive counterexamples. Subjective theories move toward the middle by imposing ever more value-laden requirements on their basic attitudinal component of pleasure or desire-fulfillment. Meanwhile, objective theories shift in the opposite direction by adding the requirement of a pro-attitude, such as endorsement, to the items on their list of prudential goods.” (Sumner, 2006: 15)

The move from subjective hedonism towards a value-laden restricted version of hedonism can be seen in Feldman’s Desert Adjusted Intrinsic Attitudinal Hedonism (2002). Similarly, the move from subjective desire fulfillment to a restricted version of desire theory can be seen in Griffin’s idealized desire satisfaction (1986). If different formal theories can be shown to lead to the same content, policy may not need to commit to any one formal theory of well-being. However, lack of willingness to commit to a particular theory may weaken any subsequent ability to respond to challenges to the constituents of well-being (Crisp, 2001).

SWB is particularly interesting in part because it is ‘thin’ and it would be expected that SWB could proxy for well-being for any good theoretical account in terms of measurement, under normal circumstances.

The similarities between some versions of the different accounts of well-being suggests that at a measurement level there may be considerable convergence between some instruments drawn from the different accounts. However, measurement instruments may show considerable divergence. For example, where preference satisfaction is measured by income this would be expected to diverge more from evaluative measures than would preference satisfaction as measured by preferences towards alternative lives.

The extent of divergence between instruments is ultimately an empirical question, which will be explored further in next chapter.
Chapter Three: Empirical differences between measures of well-being

3.1 Introduction

Within the many conceptions of well-being, different measurement methods are used. This results in many different answers to the question ‘what is i’s well-being?’ How much does it matter, from a policy perspective, if a subjective measure of well-being, such as self-reported life satisfaction, is used instead of more traditional income or consumption measures? How much does it matter whether life satisfaction or happiness, or mental health, or a weighted average of satisfaction across a range of domains, is chosen as the measure of subjective well-being (SWB)? If the answer is a good deal, then policy makers should be concerned about their choice of well-being indicator and understand the implications of that choice.

This chapter will consider evidence from existing literature on the similarities between different well-being measures and the extent of divergence in the determinants of well-being arising from the use of different well-being measures. Data from the British Household Panel Survey (BHPS) will be used to explore differences both between alternative methods for measuring the same conception of well-being and measures that adopt a different conception of well-being. The BHPS data will provide an empirical example of the differences between the possible measures of well-being.

This analysis explores a range of issues. Firstly, it will consider the degree of variability between the measures by looking at simple correlations. It will consider the types of people who would register different levels of well-being according to the different measures. This will be done by comparing, through converting well-being measures into standardised z scores, the average position in the distribution (according to the different measures) of people with different characteristics. Lastly, consideration will be given to those factors which are associated with higher well-being for each of the measures when holding other factors constant, looking at both differences in levels and change across the variables.

3.2 Evidence from the literature

3.2.1 Income as a proxy of utility versus SWB

Since Easterlin (1974) and Scitovsky (1976) economists have questioned the link between economic growth and improvements in subjective indicators of well-being. These doubts drew
upon the lack of a strong cross-country relationship between income and SWB measures, and the stable average life satisfaction during periods of significant economic growth in countries such as the US, UK and Japan (Easterlin, 1995, 2001, 2005a; Veenhoven, 1993). More recent analysis in a range of cross-country and within-country data almost always finds a significant association between income and SWB (see Clark et al. (2007) and Dolan et al. (2006a) for recent reviews) and some research suggests the relationship to be causal (Gardner and Oswald, 2006; Frijters et al. 2004). However, the association is generally weak, particularly for high-income groups, suggesting a divergence between subjective measures and income. There are also some interesting exceptions that find a negative or non-significant relationship (Clark and Oswald, 1994; Headey and Wooden, 2004; Lee et al., 2001; Magdol, 2002; McBride 2001; Ritchey et al., 2001; Shields and Wailoo, 2002; Wildman and Jones, 2002; van den Berg and Ferrer-i-Carbonell, forthcoming). Consequently, income and SWB can offer quite different answers to the question ‘what is i’s well-being?’.

3.2.2 Various subjective measures of well-being

My aim is to judge the impact of the choice of measurement instrument on (i) SWB level, and (ii) the relationship between SWB and economic and social factors (in this chapter, SWB is taken to be any mental-state account).

Whilst some comparison across findings from different studies may be useful to address this, it will be limited by the fact that the time period, and many other observed and unobserved variables may differ in addition to the choice of dependent variable. Fortunately, some studies contain more than one measure of SWB allowing direct comparisons to be made. That said, different measures often use different response scales (for example, the World Values Survey (WVS) uses a four point overall happiness question and a ten point life satisfaction question), which may contribute to differences in significant determinants in addition to the conceptual and language difference.

Reports of life satisfaction are generally found to correlate strongly with reports of overall happiness. For example, Inglehart and Klingemann (2000) report a correlation between average happiness and life satisfaction of 0.81 from the WVS and Di Tella et al. (2003) report a correlation of 0.56 in Eurobarometer data. That the correlation is not higher suggests that one or both of the scales contain measurement error, or that they measure different concepts.

Most studies find remarkably similar covariates between the single life satisfaction question and the single happiness question (e.g. Blanchflower and Oswald (2004) and Di Tella et al. (2003) for the Eurobarometer; Helliwell and Putnam (2004) for the WVS). Where differences are
found, it has been hypothesised that the life satisfaction question “triggers answers that are more reflective of one’s whole life experiences than one’s current circumstances or mood” (Helliwell and Putnam, 2004: 1438). Despite the intuitive appeal of this statement, there is minimal evidence to support it. So, more evidence is needed to determine the exact relationship between the overall happiness and life satisfaction questions.

Greater variation may be expected between evaluative measures of SWB and those focused on depression and poor mental health. However, Thoits and Hewitt (2001), using US Americans’ Changing Lives (ACL) study, found that life satisfaction, overall happiness and the CES-D depression scale were generally related to the same variables. Similarly, Helliwell (2006), using WVS data, found that suicide rates and life satisfaction were affected by the same variables. As discussed in Chapter Two, there is some evidence to suggest that positive and negative affect are different attributes, which relate to different factors and show surprisingly low correlation, however, this is not universally accepted.

Despite the overwhelming similarity of covariates across the different measures, there are a few interesting differences, particularly in relation to gender, age and education.

Controlling for a range of other factors, being female tends to worsen mental health measures but increase, or not relate to, life satisfaction scores (Baker et al. (2005) and Ritchey et al. (2001) using the ACL study, and Helliwell and Putman (2004) using the WVS). Other studies tend to find that women report higher scores both for positive measures (e.g. happiness, Alesina et al., 2004) and negative measures of well-being (e.g. CES-D scores, Kim and McKenry, 2002; GHQ scores in the BHPS, Clark and Oswald, 1994). However, it is unclear whether this greater range is due to greater variance in actual emotional experiences or greater willingness to report emotional diversity.

The relationship between age and SWB also shows some differences across measures. Age is positively related to happiness, but not life satisfaction or depression (CES-D) in the ACL study (Baker et al., 2005). Although using the same study Ritchey et al. (2001) find that age is positively related to life satisfaction and overall happiness but not depression. Helliwell and Putnam (2004) find that those over 55 are less happy than those under 25, but not less satisfied with their lives. Using the US MIDUS study Greenfield and Marks (2004) find that age reduces negative affect, but is not significantly related to positive affect or purpose in life. With the same study Keyes et al. (2002) find that age is more strongly correlated to a SWB measure than Ryff’s Psychological Well-being (PWB) measure. Ryff (1989) in a small cross-section sample found that older respondents had similar affect balance and self-esteem but lower purpose in life.
and lower personal growth. However, some of these studies do not account for non-linearities in the relationship between age and SWB.

The prevalence of clinical depression is generally lower in old age (Jorm, 2000; Ernst and Angst, 1995), which would be consistent with a positive – although non-linear – relationship usually found between age and SWB indicators. Potentially the relationship between age and outcome measure may vary depending on the extent to which the measure draws upon activity occurring within a person's life (e.g. satisfaction with leisure time), or contemplative and contentment aspects. However, as yet the evidence is insufficient to offer a clear conclusion.

Education variables can show quite dramatic differences between the measures. Baker et al. (2005), Greenfield and Marks (2004), Thoits and Hewitt (2001) and Ritchey et al. (2001) all analyse the ACL study to find that education has a negative effect on life satisfaction, either reduces or is not significantly related to overall happiness and positive affect, either reduces or has an insignificant effect on depression (CES-D), and increases purpose in life. Similarly, using the WVS, Putnam and Helliwell (2004) find that education reduces life satisfaction, but is not significantly related to overall happiness. Keyes et al. (2002) using the US MIDUS study find that the correlation between life satisfaction and education is lower than the correlation between Ryff’s PWB and education (0.07 vs 0.2). This may be linked to the inclusion of control and autonomy within the PWB. Borooah (2005), in a study in Belfast, found that high education reduces the probability of being unhappy but does not affect the probability of using tranquillizers or having thoughts of self-harm.

The findings on education are particularly interesting since they do not rest on findings of significance compared with non-significance but rather, at times, significant differences in the direction of the effect. The negative effect of education appears to be limited to life satisfaction. This could arise if the life satisfaction measure called upon a more comparative type assessment in which expectations, which are likely to be raised through education, come more into play. Clark and Oswald (1996) suggest that their finding of a negative relationship between education and job satisfaction in the BHPS may be explained if education raises aspirations, and satisfaction depends upon the gap between outcomes and aspiration.22

---

22 They also note that this relationship may be caused by the drop in relative income of those with qualifications compared to those without following the recession in Britain in the early 1990s. The potentially comparative nature of life satisfaction and overall happiness measures may also contribute to the finding of a significant relative income effect when using these measures (Blanchflower and Oswald, 2004; Dorn et al., 2007; Ferrer-i-Carbonell, 2005; Luttmer, 2005; Hudson, 2006; Weinzierl, 2005) but a non-significant effect when using the GQH (Clark, 2003b).
Understanding the relationship between well-being and education is more policy relevant than that of gender and age because education is a target of policy. The literature suggests a divergence in this relationship according to well-being measure, therefore understanding this further will be a focus of the data analysis.

### 3.3 Well-being measures available in the BHPS

The BHPS has a wealth of possible well-being indicators that are described in detail below. Preference satisfaction measures include income and consumption, plus a preference-based health-related quality of life (HRQoL) measure.

The BHPS also contains a range of questions that would be of interest to substantive good theorists (discussed in Appendix III.1.3). Anand et al., (2005) have used 28 questions from the BHPS to pick up various capabilities. These are chosen to be compatible with the list of essential human functionings identified by Nussbaum (2000, see Appendix III). The authors regress these capabilities against overall life satisfaction, thereby assuming substitutability between capabilities. They note that “some capabilities have a bigger impact on well-being than others” (Anand et al., 2005: 42), which implies an identification of well-being with life satisfaction.

The measures that I investigate here are single index measures that are empirically comparable. Whilst it would be interesting to include a measure from a substantive good perspective, particularly one which focuses on capabilities since this is of growing interest in policy circles (Burchardt and Vizard, 2007), to do so faces some difficulties.

Firstly, if we are willing to trade-off one capability for another we need to know the weights assigned to each capability. If weights are based on predicted life satisfaction this undermines the unique advantage (not having to be concerned with the subjects own view of their life) of adopting a capabilities perspective. The choice of weighting system (e.g. implicitly through number of responses within each domain and response options, or explicitly by weighting each domain equally) will impact upon findings. For example, if education is included as a capability then the weight given to education attainment will influence the strength of any relationship between education and well-being as defined by capabilities. The same would apply to income or health. Yet knowing how education, income or health impact upon opportunities for living a good live are important, and seemingly open, questions.

A second problem arises through the incomplete nature of capabilities within the BHPS. Anand et al.’s 28 questions (2005) do not capture the dimensions of ‘life’ and ‘living in harmony with other species’ which form part of Nussbaum’s list, and in many cases, due to data limitations,
the questions do not provide a good match for the capability. Indeed, without the constraints of the BHPS, 65 questions were used (Anand et al., undated). If an incomplete measure of capability is used it will be impossible to distinguish differences between well-being measures which arise due to the incompleteness of the measure, and those arising from conceptual differences. For this exercise inclusion of a capabilities measure will raise more questions than it will answer. However, empirical work on measuring capabilities and flourishing measures is an important area for future development.

Evaluative well-being is captured through the life satisfaction questions, and a weighted aggregate of domain satisfactions. Both these are seen as SWB measures. SWB can also be captured through the measure of psychological health, the GHQ, and the hybrid measure, the CASP-19.

Each of these measures will be discussed in more detail below.

### 3.3.1 Income and consumption

The main measurement tool for preference satisfaction is income. The income data collected in the BHPS allows income to be presented in a range of different ways: individual or household; gross or net; annual or current; equivalised or non-equivalised. However, in most cases there is no strong theoretical claim to a particular income measure as being the measure which most closely represents utility or satisfaction of individual preferences.

Since only 15 percent of households in the 2004 sample are single occupants it would appear that, even if all households do not share income, for a majority some sharing of income is likely to occur, suggesting a need to consider household in addition to individual income.

Gross and net household income are available, the latter as a supplementary data file (Bardasi and Jenkins, 2004; Levy et al., 2006)\(^23\). From a welfare economics perspective, net, rather than gross, income will align most closely with preference satisfaction. If the relationship between net and gross income were similar between individuals it would matter little which was chosen. However, the deviation between gross and net income is potentially linked to other important variables, such as employment status (for example students are not required to pay council tax).

\(^{23}\) Household net current income is the sum across all household members of: cash income from all sources (income from employment and self-employment, investments and savings, private and occupational pensions, and other market income, plus cash social security and social assistance receipts and private transfers (e.g. maintenance) minus direct taxes (income tax, employee National Insurance Contributions, local taxes such as the community charge and the council tax) and occupational pension contributions. Net income excludes some cases where data on all individuals in the household is not available.
Furthermore, under a progressive taxation system net incomes should narrow the distribution of income, resulting in the difference between net and gross income being dependent upon income.

Both current monthly/weekly income and annual income are available. These are derived slightly differently\(^{24}\), and are subject to different sources of measurement and recall error.

Rather than focusing on current or annual income, a measure of permanent income may give a better guide to the individual’s ability to meet her preferences. Permanent income would include all income, assets and debts, plus depredicted future income streams. Estimating permanent income is very data intensive (even with the wealth of data available in the BHPS this would be problematic, for example, full details of savings and investments are only available in two waves). Differences in methodology (for example different methods for valuing assets and predicting future income) would result in considerable potential variability in the measure. Current consumption is usually taken as a better predictor of permanent income than current income since consumption can be smoothed by saving and dis-saving.

In the BHPS consumption data is only included for certain items. These include household food\(^{25}\), energy costs, housing costs (rent or mortgage, after any rent rebate), childcare, private medical insurance, meals out, leisure expenditure, and some durable goods. In 2004/5 (wave 14) on average 60% of household expenditure reported was attributed to food and 34% on housing and fuel. Household consumption on food in the Expenditure and Food Survey (ONS) for 2004/5 represented 10.3% of the total consumption, and housing and fuel costs were 9.2% which would imply that about one sixth to one third of all household consumption is represented by this BHPS consumption data, depending on which is used as a reference point. However, depending on age, family circumstances, employment etc. the percentage of total expenditure on these items will vary. For example, those in the bottom 10% of the income distribution (gross, household) spend 16.0% of total expenditure on food, whereas those in the top 10% of the income distribution spend 7.4% of total expenditure on food, as would be expected by Engel’s Law (ONS, 2005, Table A6). This would suggest that these items underestimate total consumption for households higher up the income distribution. Whilst current consumption may give a better indication of permanent income than current income, the relationship between consumption and permanent income will still depend upon position in the life cycle, and may also change over time with changes in the business cycle and trends in savings rates.

---

\(^{24}\) For example, annual net household income is similar to current net household income variable except local taxes have not been deducted from income, and earnings from a second job are also excluded.  
\(^{25}\) This is reported in 12 categorise since wave two, hence a mid point of each category is taken.
To allow for some sharing of income at the household level, it is necessary to allocate household income to individuals within the household. As discussed in Chapter One, this can be done by equivalising income. Analysis here is restricted to the McClements and the modified-OECD scales since they are most commonly used in the UK.

The above discussion shows that there is no unique, standardised measure for income that can be used to represent individual preference satisfaction. Critically, the extent of the variation between the different income measures is considerable. A simple correlation between these different measures in the BHPS data for wave 14 shows this variability (see Appendix 3.A). The consumption measures are not strongly correlated with the income measures, which may reflect limitations with the consumption data in the BHPS. Despite the potential for differences between the equivalence scales, equivalised income by McClements and the modified-OECD scales are highly correlated. Net income is highly correlated to gross income measures, however, as noted above there may still be an important minority of cases where these differ. Individual and household income are not very strongly correlated (0.62 for annual gross, 0.65 for current gross), this may not be surprising, but does point to the importance of understanding intra-household distributions of income if income is to be used as a measure of individual well-being.

The divergence between income measures can also be reflected in the difference between rank position according to the different income measures. If all income measures are considered, the average range of rank percentile for wave 14 is 31% points. Considering household income measures only, this figure is 19% points. Consequently, the choice of income measure makes a considerable difference to the level of well-being attributed to each individual.

In order to make the comparisons with other well-being measures, without becoming overwhelmed, I have chosen one measure of income. That measure is net, current, household income equivalised using the modified-OECD scale.

The modified-OECD scale is chosen because it is becoming more popular than the McClements scale (see DWP, 2005). Current income is chosen over annual, because other well-being indicators (e.g. the GHQ and the CASP-19) refer to life in recent weeks or, in the case of life satisfaction, the last year from the interview, rather than the data collection year, September to

For example, a single mother with a child under 2 would receive a weight of 0.7 of an adult couple household within the McClements scale (giving an equivalent household income for £20,000 of £28,571) whereas they would receive a weight of 0.87 within the modified OECD scale (giving an equivalent income of £22,989).
September, used by the annual income measure.\textsuperscript{27} Net income is preferred over gross as it is closer to disposable household income and consequently likely to be closer to the extent of material preferences satisfied. Household consumption data are also considered, equivalised using the modified-OECD scale for the adjustment of household size, although limitations with this data mean it may not be an accurate measure of consumption.

The simple histograms, shown in Appendix 3.C, reveal that, as expected, the income and consumption data are highly positively skewed. If income is treated as a cardinal measure of well-being, the strong positive skew would suggest a minority of very wealthy individuals having a dramatically highly level of well-being than the majority. Income and consumption is more unequal than the other well-being measures. However, this is unsurprising given that both income and consumption are non-bounded whereas the other well-being measures use bounded scales. As discussed in Chapter One, there is widespread acceptance of the diminishing marginal utility of income yet little agreement about the extent to which the utility from additional income diminishes as income levels rise. For consistency across economic evaluations within the UK the Treasury recommends the use of a constant value of elasticity of marginal utility of one, which is the equivalent of log of income. The distribution becomes close to normal when income is logged.\textsuperscript{28}

3.3.2 Health-related quality of life

Chapter Two discussed the use of QALYs as a measure of utility used in within health care resource allocation decisions. In waves 9 and 14, all respondents were asked 36 general health questions comprising the short form medical health outcomes survey, Sf-36 (Ware and Sherbourne, 1992). These responses can be combined with UK public preference data to generate values for the Sf-6d that lie on a scale from 0 (equivalent to death) to 1 (full health) (Brazier \textit{et al.}, 2002). The Sf-6d therefore provides the ‘quality’ adjustment that is used in the generation of QALYs.

3.3.3 CASP-19

The CASP-19 (Hyde \textit{et al.}, 2003), which is included in wave 11, was described in Chapter Two and is shown in detail in Appendix 3.B. CASP-19 is currently being used by the DWP as an

\textsuperscript{27} Additionally, there were some irregularities within the annual income measures in wave 12 of the dataset.
\textsuperscript{28} Excluding cases with negative or non-computable log of equivalised income.
indicator of the overall SWB for older people within the UK Governments Opportunity Age strategy.  

3.3.4 SWB past and future

CASP-19 contains some general questions that are an overall reflection of life. Although these specific questions have not been subject to the same psychometric testing, they present an opportunity to make additional comparisons, and incorporate reflection over the whole of life without the use of the term ‘satisfaction’. This is useful because the term satisfaction may be problematic. For example, for some it may suggest a judgement given their initial life circumstances, for others it may be indicative of a complaining attitude if responded to negatively.

Responses to the questions ‘on balance, I look back on my life with a sense of happiness’ and ‘I feel satisfied with the way my life has turned out’, combine to make the variable ‘SWB past’. Each question has a 0 to 3 response scale, giving an aggregate score of 0 to 6. ‘SWB future’ is comprised of responses to the questions ‘I look forward to each day’ and ‘I feel that the future looks good for me’.

The histograms in Appendix 3.C show that a greater percentage of respondents use the top end of the scale for both ‘past SWB’ and ‘future SWB’ than for the life satisfaction question.

3.3.5 Life satisfaction measures

From wave 7 (excluding wave 11) respondents have been asked ‘how dissatisfied or satisfied you are with your life overall’ using a response scale from 1 (not satisfied) to 7 (completely satisfied). Most people respond above the mid point, the modal response being six. Respondents are also asked eight domain satisfactions (health, income of the household, house/flat, spouse/partner, job, social life, amount of leisure time and use of leisure time) using the same response scale. Table 1 shows the mean and standard deviation for each domain satisfaction.

<table>
<thead>
<tr>
<th>Domain satisfactions, wave 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Life overall</td>
</tr>
<tr>
<td>Health</td>
</tr>
<tr>
<td>Income of the household</td>
</tr>
<tr>
<td>House/flat</td>
</tr>
<tr>
<td>Spouse/partner</td>
</tr>
<tr>
<td>Job</td>
</tr>
</tbody>
</table>

29 http://www.dwp.gov.uk/opportunity_age/indicators/indicators_table.asp
Social life  4.981  1.325  
Amount of leisure time  4.473  1.476  
Use of leisure time   4.802  1.384  

Note: Those who respond to all satisfaction questions, excluding Northern Ireland. Number of observations: 5,988

Domain satisfactions are important in their own right, and considerable work has explored the determinants of job satisfaction (Freeman, 1978; Clark and Oswald, 1996; Clark, 1999, Diaz-Serrano and Cabral Vieira, 2005), housing satisfaction (Diaz-Serrano, 2006) and income satisfaction (Bellemare et al., 2002; Ferrer-i-Carbonell and Van Praag, 2003). However, the aim here is to combine this information to generate a single well-being index. As discussed in Chapter Two, weights need to be attributed to the different domains in order to generate a single index representing a general assessment of life. Chapter Two concluded that the best means for deriving weights is to regress the domain satisfactions on the single life satisfaction question. This rests on the assumption that the life satisfaction response represents true well-being, measured with some error. A weighted aggregation of domain satisfaction, because it is drawn from responses to eight questions rather than just one, offers the potential for a more sensitive measure of well-being, with potentially reduced measurement error.

To generate the weights for the BHPS, life satisfaction is treated as a weighted average of satisfaction in each of the eight domains, along with fixed time effects. The model is as follows:

$$LS_t = \delta_1 DS_{1t} + \ldots + \delta_8 DS_{8t} + \beta Year_t + \varepsilon_u$$  \hspace{1cm} (1)

(i = 1…n) where n is the number of individuals  
(t = 1…T) where T is the number of time periods  
where ‘Year’ represents the fixed time effects and the error term $\varepsilon_u$ has two components

$$\varepsilon_u = \alpha_i + \nu_u$$

where $\alpha_i$ is an individual, time invariant effect. The individual effect ($\alpha_i$) in this model, which can be thought of as something like an individual’s genetic happiness level, shifts the level of well-being; a happy disposition, or tendency to answer survey questions in a favourable manner, will shift the level of well-being upwards in each time period. In order to get estimates of the domain weights ($\delta_1 \ldots \delta_8$), it is necessary either to estimate the unobserved, individual effect or remove it from the model. If an additional assumption is made that the individual effect is not correlated with the independent variables then it can be treated as a random variable with a
probability distribution $E(\alpha_i) = 0$. Under these circumstances a Generalised Least Squares\(^{30}\) (random effects model) will give the most efficient estimates of $\delta_1,...,\delta_k$. These will be a weighted average of the estimators generated by looking at only within person variation and those generated by looking only at between person variation.

The use of GLS, random effects (re), will result in biased and inconsistent estimates of the coefficients if an incorrect assumption has been made about the distribution of the individual effects ($\alpha_i$). If there is some correlation between the unobserved, individual effect and the independent variables, then we can expect coefficient estimates to be bias. For example, if it is the case that generally optimistic people are likely to respond more positively to say both life satisfaction and satisfaction with their job the estimate for the coefficient on job satisfaction is likely to be an overestimate.

Rather than estimating the individual effect, as done under GLS, the fixed effects (fe) approach uses the deviation from the individual mean for each variable, as in (2), to remove the individual effects.

$$\text{LS}_i - \overline{\text{LS}}_i = \delta (DS_{ui} - \overline{DS}_i) + (\mu_i - \overline{\mu}) + (\epsilon_i - \overline{\epsilon}) \quad (2)$$

Since the aim here is to provide the best prediction of life satisfaction rather than consider any relationship between separate domain satisfactions and overall satisfaction, estimates using both the within and between effects may be preferred due to their greater explanatory power. The GLS re models have an overall R squared of around 0.56 to 0.60 whereas the OLS fe models have a within R squared of around 0.35 (see Table 2, below). Although the coefficient sizes and relativities change very little when tested with a fe model which controls for the individual effect by considering only the within person variation (indeed, a satisfaction index generated from fe weights correlates 0.999 with a satisfaction index from the re weights), the null hypothesis of no differences between the coefficients is rejected by the Hausman test\(^{31}\), suggesting superiority of the fe model.

There may be concern over omitted variables that may also be correlated with domain satisfactions, suggesting the true model would be:

---

\(^{30}\) The composite error term exhibits a particular form of autocorrelation (unless $\sigma^2 = 0$) which would make standard errors for OLS incorrect. A more efficient GLS estimator exploits the structure of the error covariance matrix (Verbeek, 2000: 348).

\(^{31}\) A Hausman test (1978) compares the fe and re estimators. Under the null hypothesis that $\alpha_i$ and $X_{ui}$ are not correlated these two estimators will not be significantly different, hence both estimators would be consistent but the re estimator would be efficient.
Chapter 3  
Empirical differences between measures

\[ LS_{it} = \delta_1 DS_{it} + ... \delta_8 DS_{8it} + \beta Year_{it} + X_{it} + \varepsilon_{it} \]  
(3)

The omitted variable will be picked up within the error, which if correlated with any of the domain satisfactions will result in endogeneity bias. For example, age may have a direct effect on life satisfaction independent of any indirect effect via domain satisfactions. If age is correlated with any of the domain satisfactions, say for example satisfaction with housing, the coefficient on this domain satisfaction will be inconsistent. This bias would mean changes in satisfaction with housing would not then result in the anticipated changes in life satisfaction. However, if weights are derived from a model including possible variables which impact upon life satisfaction directly, domain satisfactions will not represent a complete picture of domain-weighted life satisfaction. Either additional personal characteristics will be necessary, which implies considerable data demands, or a prediction using only the domain weights from such a model will systematically under or estimate overall satisfaction according to additional characteristics included. For example, if age increases life satisfaction, domain weights not accounting for this will underestimate life satisfaction for older people.

Including additional control variables into the model gives similar domain weights, as can be seen in Table 2. The additional variables also add little to the explanatory power of the model. It is therefore reasonable to use the weights from the initial model (1).

The life and domain satisfaction questions are strictly ordinal, however, this method treats them as cardinal. To generate weights for the domains, whilst respecting the ordinality of the data, would require using a random effects ordered probit model and including six dummies for each domain satisfaction (giving 48 dummies). This would then make assigning weights to the domain satisfactions somewhat complex. Ferrer-i-Carbonell (2003) regressed domain satisfactions against a single life satisfaction using the German data (GSOEP) and showed that using GLS or ordered probit made little difference to the relative trade-offs between the variables (Ferrer-i-Carbonell, 2003: 49). Although her model included domain satisfactions as cardinal variables rather than using (in her case) 60 dummy variables, it does offer some confidence in the assumption of cardinality for life satisfaction.

Preserving the ordinality of the data whilst simultaneously addressing the unobserved heterogeneity can be achieved by direct modelling of the unobserved individual effect through controlling for the individual mean level of the independent variables via a “Mundlak”
approach\(^{32}\) (Mundlak, 1978). This approach would add considerably to the complexity and may still not adequately control for the individual effect.

Another option is to convert the life satisfaction response into a dichotomous variable, ‘LShigh’ which is true whenever an individual scores above a cut off (such as the mean) for life satisfaction. This can be modelled as an underlying latent variable;

\[
LS^*_{it} = \delta DS_{1it} + ... + \delta_k DS_{kit} + \beta Year_{it} + \varepsilon_{it}
\]

where we observe \(LShigh_{it} = 1\) if \(LS^*_{it} > 0\) and \(LShigh_{it} = 0\) otherwise.

\[
P(LShigh_{it} = 1; \delta DS_i \beta \alpha) = \frac{\exp(\delta DS_{it} + \beta Year_{it} + \alpha_i)}{1 + \exp(\delta DS_{it} + \beta Year_{it} + \alpha_i)}
\]

where \(G(.)\) can be the probit or logit function used to model the probability of LShigh being 1 for a given individual in a given time period, conditional on the characteristics in time period \(t\) for individual \(i\) and the unobserved, time invariant, individual effect \((\alpha_i)\), as in (5).\(^{33}\)

Chamberlain (1980) has shown that it is possible to use Maximum Likelihood estimation, control for fixed effects and get consistent estimates if the factor likelihood is divided into two, one part which is conditional on incidental parameters (all the individual effects, which increase with \(N\)), and a part which does not depend on the individual effects. This requires discarding individuals who show no change in state during the time period.

This method therefore involves considerable loss of information, partly due to the dichotomisation of the 1-7 scale, and partly because it only incorporates individuals who move across the dichotomised scale during the time period. Ferrer-i-Carbonell and Frijters (2004) extend this method to generate a dependent variable which is dichotomised at the individual level, therefore where there is any individual change across the time period this change is reflected, hence giving a much greater sample size than in the conditional fe logit case which uses the same cut off point for all individuals. However, this method still results in a considerable loss of information and is yet to be widely replicated. Ferrer-i-Carbonell and Frijters (2004) also show that controlling adequately for the unobserved individual effects will have a greater impact upon determinants of SWB compared to respecting the ordinality of the

\(^{32}\) van Praag et al. (2003) use a Mundlak approach and disaggregate the impact of six domain satisfactions available within the GSOEP onto general satisfaction into shock effects and level effects.

\(^{33}\) Using a Maximum Likelihood, with respect to \(\beta\) and \(\alpha\) when \(T\) is not approaching infinity would lead to inconsistent estimates since for a fixed \(T\) the number of parameters grows with the size of \(N\) (as would including \(N\) dummy variables to account for the fixed effect \(\alpha_i\)) (Neyman and Scott, 1948). The incidental parameters problems can be overcome (hence giving a consistent estimate for \(\alpha\) and \(\beta\)) if there are a growing number of observations for each individual.
life satisfaction data. Consequently, to make best use of the data, control for the individual fixed effects and for ease of interpretation, the use of OLS fixed effects is preferred here.

Table 2: Coefficients on domain satisfactions from GLS-RE & OLS-FE on life satisfaction

<table>
<thead>
<tr>
<th>Satisfaction with…</th>
<th>RE</th>
<th>FE</th>
<th>FE Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>0.124</td>
<td>0.108</td>
<td>0.099</td>
</tr>
<tr>
<td>Income</td>
<td>0.081</td>
<td>0.076</td>
<td>0.076</td>
</tr>
<tr>
<td>House/flat</td>
<td>0.057</td>
<td>0.052</td>
<td>0.051</td>
</tr>
<tr>
<td>Partner</td>
<td>0.191</td>
<td>0.186</td>
<td>0.185</td>
</tr>
<tr>
<td>Job</td>
<td>0.114</td>
<td>0.109</td>
<td>0.107</td>
</tr>
<tr>
<td>Social life</td>
<td>0.140</td>
<td>0.123</td>
<td>0.126</td>
</tr>
<tr>
<td>Amount leisure time</td>
<td>0.041</td>
<td>0.045</td>
<td>0.047</td>
</tr>
<tr>
<td>Use leisure time</td>
<td>0.135</td>
<td>0.125</td>
<td>0.125</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Other controls</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>R squared overall</td>
<td>0.5685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R squared within</td>
<td></td>
<td>0.3505</td>
<td>0.3543</td>
</tr>
<tr>
<td>Observations</td>
<td>49970</td>
<td>49970</td>
<td>49368</td>
</tr>
<tr>
<td>Number of groups</td>
<td>11452</td>
<td>11452</td>
<td>11363</td>
</tr>
</tbody>
</table>

Wald chi²(16) = 42501
F(16, 38502) = 1298
F(54, 37951) = 385
P>chi² 0.000
P>F = 0.000

Note: Waves 6 to 15, excluding 11. For those who report satisfaction with their job and their partner. P<0.000 for all domains. The Breusch Pagan test of individual effects confirms that the data cannot appropriately be treated as pooled data. Other controls include age dummies, job status, health status, household composition (children and adults), high load carer, and the log of household gross current income.

The domains in the BHPS include satisfaction with one’s partner and with one’s job, to which many respondents answer ‘not relevant’. It is therefore necessary to calculate four sets of weights; those who reply to all domain questions, those who reply to the job question but not partner question, those who reply to partner question but not job question and those who reply to neither.

The R squared from the random effects model implies that around 50-60% of the variance in life satisfaction can be explained by the domain satisfactions. This suggests that considerable variability in life satisfaction is not picked up by these domains satisfactions (although this compares favourably to the PWI which predicts 30-60% of life satisfaction variance for a range of different country datasets (International Well-being group, 2006)). This may be the result of measurement error, either in the domain responses or the life satisfaction response, or the exclusion of important domains. Compared to the domains in the PWI the BHPS selection excludes satisfaction with achievements in life, safety, future security, spirituality or religion, feeling part of one’s community, and for those without a partner, personal relationships.

Breusch-Pagan test is a Lagrange multiplier test with a null hypothesis that there are no individual specific effects, hence \( \sigma^2 = 0 \) (Greene, 2000: 298).
Consequently, we would expect considerable measurement error in the domain-weighted satisfaction measure as important domains which are intuitively connected with well-being have been excluded.

Variation between domain-weighted and the single life satisfaction response may also arise if the global satisfaction measure is influenced more heavily than domain satisfactions by current mood (Schwartz and Strack, 1999), by desires for self-affirmation (Haybron, 2005) or by what is described by Cummins and Nistico (2002) as the ‘homeostatic’ system. Within a ‘homeostatic’ system SWB is seen as being actively controlled and maintained by a set of psychological devices that operate to a lesser extent on the more specific domain satisfactions.

Even if an agreed upon set of domains is determined and any redundancy removed through the use of factor analysis, generating a single weighted index still rests on the assumption that it is appropriate to apply the same weights to all individuals. This assumption faces concerns that weights given to the domains could change over time, and that weights may vary in systematic ways across individuals, for example by age, gender, income, and family circumstances.

Table 3 shows some variability between weights derived from applying (1) to various subgroups; gender, those under 40 and over 60, those on high household incomes (above £38,400 in 2005 prices), and those on low household incomes (below £19,200). Across all groups satisfaction with the amount of leisure time has the least impact upon overall happiness, and the three most heavily weighted domains are partner, social life and the use of leisure time, underlying the importance of intimate and social relationships.

Table 3: Coefficients on domain satisfactions from an OLS-FE on life satisfaction, by subgroup

<table>
<thead>
<tr>
<th>Satisfaction with…</th>
<th>Males</th>
<th>Females</th>
<th>Over 60</th>
<th>Under 40</th>
<th>High income</th>
<th>Low income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>0.098</td>
<td>0.118</td>
<td>0.070</td>
<td>0.116</td>
<td>0.108</td>
<td>0.106</td>
</tr>
<tr>
<td>Income</td>
<td>0.075</td>
<td>0.072</td>
<td>0.029</td>
<td>0.082</td>
<td>0.077</td>
<td>0.066</td>
</tr>
<tr>
<td>House/flat</td>
<td>0.047</td>
<td>0.057</td>
<td>0.063</td>
<td>0.049</td>
<td>0.048</td>
<td>0.054</td>
</tr>
<tr>
<td>Partner</td>
<td>0.165</td>
<td>0.207</td>
<td>0.132</td>
<td>0.196</td>
<td>0.188</td>
<td>0.149</td>
</tr>
<tr>
<td>Job</td>
<td>0.133</td>
<td>0.087</td>
<td>0.121</td>
<td>0.113</td>
<td>0.127</td>
<td>0.084</td>
</tr>
<tr>
<td>Social life</td>
<td>0.130</td>
<td>0.116</td>
<td>0.172</td>
<td>0.101</td>
<td>0.110</td>
<td>0.150</td>
</tr>
<tr>
<td>Amount leisure time</td>
<td>0.041</td>
<td>0.048</td>
<td>0.050</td>
<td>0.037</td>
<td>0.055</td>
<td>0.039</td>
</tr>
<tr>
<td>Use leisure time</td>
<td>0.124</td>
<td>0.125</td>
<td>0.144</td>
<td>0.117</td>
<td>0.125</td>
<td>0.163</td>
</tr>
</tbody>
</table>

Note: Those with a relationship and a job, waves 7 to 15 (excluding 11). Time dummies included. All domains $P<0.000$. Despite slight differences in weights, particularly between the older and younger subgroup, the domain-weighted satisfaction variable for each subgroup is highly correlated (>0.99) to the
population-weighted variable. This suggests there is minimal loss of accuracy from using population-based weights.

The aim of generating a domain-satisfaction index from a prediction of life satisfaction has been met. However, two outstanding problems remain.

Firstly, as discussed in detail in Chapter Two, this approach also assumes a “bottom-up” construction of overall life satisfaction, yet there may be some reverse causality operating. This is important to consider from a policy perspective since the domain-weighted satisfaction approach is likely to be appealing to policy makers whose remit fits a particular domain. To the extent that the causality runs from life satisfaction to domain satisfaction, even if these domain weights provide an accurate prediction life satisfaction, changes in domain satisfaction will not necessarily result in subsequent change in life satisfaction. Indeed, changes in domain satisfaction may not be possible without a general change in the overall level of life satisfaction.

Secondly, the model used here assumes that life satisfaction is a linear combination of domain satisfaction, with constant returns to scale and constant substitutability. This implies that an individual could compensate for low satisfaction in one area of their life by improved satisfaction in another. More research is needed to confirm the accuracy of these assumptions.

### 3.3.6 Depression and anxiety scale GHQ

The General Health Questionnaire 12 (GHQ-12), found in all waves, was discussed briefly in Chapter Two and can be seen in full in Appendix 3.B. The GHQ-12 includes six positive and six negative states and a choice of four options for each in which the presence or intensity of the state over the last few weeks is related to its usual frequency or intensity, thereby creating a 36 point ‘Likert’ scale. Alternatively, the number of questions to which the individual responds in the worse two categories may be aggregated to give a 12-point ‘Caseness’ score. Looking at all 15 waves, the two GHQ scores correlate at 0.900 (obs = 159,489), although high, this still indicates some deviation between the two scoring methods.

As discussed in Chapter Two, the simple aggregation of the positive and negative responses within affect-based scales, including the GHQ, has received criticism by those who see these domains as separate. Consequently, I introduce an additional scoring system, in which the six positive questions are aggregated, giving the ‘GHQ-positive’, and the six negative questions are aggregated giving the ‘GHQ-negative’, using the ‘Likert’ style scoring. Across the 15 waves
these only correlate at 0.636 (observations: 159,484)\(^{35}\), suggesting a potential divergence between these measures. However, the positive and negative questions of the GHQ (see Appendix 3.B) do not closely match positive and negative affect. Both the positive questions and negative questions include aspects of feeling, functionings and evaluation. This suggests the scale identifies attributes beyond positive and negative affect.

The ‘Likert’ and ‘Caseness’ scoring methods for the GHQ-12 show a negative skew in the distribution. The distribution of the GHQ ‘Caseness’ score suggests it is not suited to distinguishing at the high well-being end of the scale, with over 55% of individuals scoring as having no psychological problems. The GHQ-negative and GHQ-positive show quite different distributions, the latter being more symmetrical.

### 3.4 Variability between different well-being measures in the BHPS

As seen in Chapter Two, different measures of well-being draw from different conceptions of well-being. Income and consumption, and to some extent the Sf-6d, rely upon a preference satisfaction account of well-being, in which life goes better for the individual when they have more of their preferences or desires met.

The other measures rely upon a subjective account of well-being, in which life goes better for the individual when they think and feel more positively about their life.

The CASP-19 is a subjective measure that derives from a need-based perspective of well-being and incorporates the individual’s subjective perception of whether they meet certain functionings and achievements. Some questions appear close to a subjective evaluation of life, for example, “I feel satisfied with the way my life has turned out” but others are closer to externally determined functionings, for example, “I choose to do things that I have never done before” or “family responsibilities prevent me from doing things I want to do”.

The life satisfaction and domain-weighted satisfaction measures are also compatible with an evaluative framework in which well-being is identified with an individual’s assessment of their life overall. In the single life satisfaction question respondents are left entirely to their own judgement to assess the quality of their life. However, for the domain-weighted satisfaction measure the set of domains is pre-determined. The GHQ can also be thought of as tapping into overall evaluation, in which certain externally determined feelings and considerations are identified as a proxy for the individual’s overall assessment.

\(^{35}\) Four observations have missing values for some of the GHQ questions but are given GHQ Caseness and Likert scores based on non-missing responses.
The subjective measures also vary in the emphasis they give to cognitive appraisals compared with mood and affect. For example, the GHQ taps more heavily into current affect and mood (suggesting closer alignment to a hedonist account of well-being), whereas global and domain satisfaction questions are more evaluative.

The measures also adopt different time frames, for example, the GHQ focuses on the last few weeks whereas the life and domain satisfaction questions ask about the previous year. Different measures also vary in the extent of focus on health, the Sf-6d focusing on entirely on physical and mental health and the GHQ focusing on mental health.

Unfortunately, there is no wave in the BHPS where all these measures are available. Comparisons are therefore made using wave 11 (which contains income, consumption, GHQ, CASP-19, SWB future, and SWB past) and wave 14 (which contains income, consumption, GHQ, life satisfaction, domain-weighted satisfaction and the Sf-6d).

Table 8 and Table 9 (Appendix 3.A) show the correlations between the various well-being measures. Consumption and income (i.e. measures of preference satisfaction) are not as strongly correlated as might have been expected, which suggests that people have quite different savings and consumption rates. However, limitations with the consumption data restrict the potential inference that can be made from such analysis. As expected, the SWB measures correlate only weakly with equivalised income and consumption. The Sf-6d is notably more correlated, suggesting a stronger relationship between health and income. The CASP-19, and sub-questions on ‘past SWB’ and ‘future SWB’ are slightly more correlated with income and consumption than the GHQ or life satisfaction.

The Sf-6d measure correlates better with the GHQ (the psychological measure of health) than life satisfaction, which is expected since both are oriented towards an evaluation of health states rather than global life evaluations.

The subjective measures (GHQ, CASP-19 and life satisfaction) correlate relatively well with each other, suggesting that they are tapping into similar underlying constructs. The past and future SWB questions correlate at 0.64 suggesting a slightly different concept is being picked up by these two measures.

36 At the beginning of the satisfaction questions the respondent is told to refer to their current situation, however, it is not known whether respondents think only about current circumstances and feelings when responding, or include their past or future.

37 This uses Pearson’s correlations that treat the variables as cardinal, however, a very similar picture is given when using Spearman’s rank correlations.
3.5 Are differences in well-being measure related to individual characteristics?

From a policy perspective it is important to know whether different kinds of people would be reported as having high well-being if one measure was used instead of another. In wave 14 no one scores in the top quintile for equivalised household income, equivalised consumption, life satisfaction, domain-weighted satisfaction and both GHQ measures. And, even more important from a policy perspective, less than 2% of respondents score in the bottom quintile for all these measures. This clearly suggests that low well-being by one measure does not imply low well-being by another.

One means of exploring the variation between different measures is to standardise each of the well-being scores then consider a subgroup of people and look at their average standardised score according to the different well-being indicators. Subgroups for age (those over 70), gender (males), education (those with a degree or equivalent), unemployment, those living in London and those who commute more than one hour a day are considered. Since this is a bivariate analysis relationships between particular characteristics may be spurious, however, the purpose here is to reveal differences between the different well-being measures rather than the subgroup characteristic and level of well-being.

As Figure 1 shows there is a dramatic difference between the average rank for income, consumption and the Sf-6d and that for life satisfaction and domain-weighted satisfaction for those respondents aged over 70. The average position for the GHQ Caseness and Likert is around the mean, however, there is a notable difference between GHQ positive and negative scores, with the over 70s scoring highly in terms of avoiding negative responses but relatively poorly for positive responses. This suggests that, at least for this subgroup, the combining of positive and negative attributes disguises some potentially interesting information.

The over 70s have average CASP-19 scores 0.2 standard deviations below the mean, falling between income, consumption and life satisfaction. This may be explained by the inclusion of functioning questions within the CASP, plus the response to ‘age preventing the individual from doing what they would like’. The ranking for ‘past SWB’ is above that for ‘future SWB’, as would be expected for this age group. Interestingly the average z score for life satisfaction is

38 Standardised to a normal distribution with mean of zero and standard deviation of one.
39 Note that the axis scales used on the graphs below vary in order to show differences between measures for each subgroup more clearly.
40 This may be an artefact of the scoring system for the GHQ, since on the positive questions consistency (same as) receives a lower score than improvement (better than).
closer to past SWB than future, which may suggest that older people use past life experiences to influence their life satisfaction responses.

**Figure 1: Mean standardised score by well-being measure for the over 70s**

![Mean standardised score by well-being measure for the over 70s](image)

Note: Wave 14 (observations: 1,069) and wave 11 (observations: 1,126).

Less variation between the well-being measures is found for men. As discussed above, the literature has found that men score slightly better on mental health measures than evaluative measures. This is somewhat born out here since they score slightly higher in the health related measures, the Sf-6d and the GHQ, than they do in life satisfaction. Again the difference between the GHQ positive and negative scores suggests that men score relatively higher in avoidance of negative attributes.
Looking at the educated subgroup, life satisfaction and domain-weighted satisfaction are less responsive to education than other well-being measures. Figure 3 shows those with a degree rank substantially higher according to income and consumption than the GHQ measures, life satisfaction or domain-weighted satisfaction. The standardised score for ‘past SWB’ and ‘future SWB’ and CASP-19 is greater than that for the satisfaction questions.

Note: Wave 14 (observations: 4,288) and Wave 11 (observations: 4,751).
Turning to the unemployed subgroup, all measures show average scores that are below the mean, although the Sf-6d is far less negative than other measures. For many subgroups the life satisfaction average standardised score falls around the mid point, which may arise if life satisfaction reports were random between individuals or insensitive to external circumstances, however, the unemployed have a considerably lower average standardised life satisfaction score, suggesting at least for employment status this is not the case.
For the registered disabled the average rankings across all well-being measures are also fairly low (around 0.3 to 0.5 standard deviations below the mean), again Sf-6d presents as something of an outlier (Figure 5). The CASP-19, gives the next lowest rank which may be because this measure incorporates aspects of functioning. That ‘past SWB’ is highest is also unsurprising, since this may refer to pre-disabled times.
Single parents also fare poorly across all well-being measures (Figure 6). Single parents and adults with children living in the household (Figure 7) have a higher rank for consumption than income. This may suggest that consumption is reflective of household needs. The health related measure, the Sf-6d, is higher than the subjective measures, which may result from the age of the subgroup. Those with children score above average on the GHQ positive scale (except for wave 14 for single parents) but below average on the GHQ negative scale, which may suggest children bring both more highs and lows. This offers further reason for disaggregating positive and negative attributes. The variation in well-being measures across those with children is interesting (although generally small), and suggests care will be needed in any conclusions linking changes in well-being to the presence of children. Domain satisfaction presents the lowest rank, those with children report lower satisfaction with income, housing, their spouse, social life, leisure time and use of leisure time.

Figure 5: Mean standardised score by well-being measure for the disabled

Note: Wave 14 (observations: 877) and wave 11 (observations: 734)
Chapter 3

Empirical differences between measures

Figure 6: Mean standardised score by well-being measure for single parents

Note: Wave 14 (observations: 517) and Wave 11 (observations: 618)

Figure 7: Mean standardised score by well-being measure for those with children
For those who live in London there is a considerable difference between their income and consumption compared to their rank on the SWB measures (Figure 8). However, this may highlight the inadequacy of income to reflect real income under large geographical price differences. The more geographically diverse group of respondents who commute for over one hour per day show a similar pattern: relatively high income and consumption but around the mean or below for the SWB measures (Figure 9).

Figure 8: Mean standardised score by well-being measure for those from London

Note: Wave 14 (observations: 3,007) and Wave 11 (observations: 3,496). Age 25-50 only.
These simple graphical demonstrations of average rank according to different well-being measure by a few subgroups highlight important differences between the well-being measures, particularly between those which proxy for preference satisfaction and the subjective measures.

### 3.6 Differences in the factors associated with well-being

Multiple regression analysis allows a consideration of factors that have a significant association with different well-being measures whilst holding other variables constant. Since not all well-being measures are available in every wave it is necessary to consider comparisons where multiple measures are available hence comparisons are made using waves 9 and 14 (pooled) and wave 11. Well-being is modelled as a function of a vector of independent variables \((Z)\) and an error term:

\[
WB_i = \beta Z_i + \epsilon_i
\]  

(4)
This does not take full advantage of the panel nature of the dataset, and where two waves are used requires adjusting standard errors to account for clustering at the individual level. Analysing single wave data in cross-section or pooled data does not model the individual heterogeneity, nor consequently any correlation between the any unobserved individual effect and the regressors. Significant relationships should therefore be treated with caution. However, this approach is sufficient for highlighting potentially interesting differences between the well-being measures.

Where data availability allows, fixed effects models can also be used. As noted above, fixed effects models ask whether changes in the well-being measures at an individual level relate to changes in the various factors. This is not possible for the CASP-19 measures since this is only asked in one year. The basic model is similar to (4) above,

\[ WB_{it} = \beta Z_{it} + \epsilon_{it} \]  

(5)

where the error structure is \( \epsilon_{it} = \alpha_i + \nu_{it} \)

For comparability only waves 9 and 14 are used because Sf-6d data is only available in those waves.\(^{41}\) The regressors used here are those found to be related to SWB in the literature (Dolan et al., 2006a). This includes income (log of household net, current income not equivalised), household composition, education (no qualifications compared with O-level or similar, A-level or similar, and higher education), health (subjective health scale, whether the respondent has problems walking, whether they are register disabled, and the number of inpatient hospital days (excluding births), marriage status (being married compared with being divorced or separated, cohabiting, never married, and widowed), job status (being employed compared with being self-employed, retired, unemployed, on long-term sick leave, a home-maker, or other job status), age (in categories), social contact (seeing neighbours and friends or family most days) and being a high-load carer (more than 50 hours per week). Full description of these variables can be found in Appendix 3.D.

Some of the well-being measures are strictly ordinal, such as life satisfaction. However, for simplicity of comparison the following analysis treats all the measures as continuous and rescaled from 0 to 1. As discussed above, within-person analysis is more easily conducted with continuous variables using ordinary least squared (OLS). Consequently, it is common within the literature to treat the life satisfaction variable as continuous (e.g. Meier and Stutzer, forthcoming; Stutzer and Frey, 2005). This simplification has been shown to make little difference to the determinants of SWB (Di Tella et al., 2001, Ferrer-i-Carbonell and Fritjers,

\(^{41}\) Consequently, a first difference model would in this case generate similar estimators.
Chapter 3

Empirical differences between measures

2004, Frey and Stutzer, 2000; Layard et al., 2007). The regression outputs from these 17 models are reported in Appendix 3.E and some of the interesting findings are presented in the main text graphically.

Income and consumption models are not included in the graphs because they are not mapped onto the 0 to 1 scale, and also because income is included as a determinant of well-being for the subjective measures, making direct comparisons problematic. In brief, equivalised income is found to be lower in presence of children in the household (also supported in the fixed effects model), for those who are divorced or separated, for high-load carers, for those in poor health, for those without employment (also supported in the fixed effects model), for those who talk to their neighbours most days (and see family and friends most days, in waves 9 & 14), and higher for those with more education qualifications. Equivalised consumption is lower in the presence of children in the household, for those in poor health, those who are not married, those who talk to their neighbours most days, those without employment (except students) and higher for those with more education. Full details are shown in Appendix 3.E Tables 1 to 3.

The unemployed, divorced and those in poor health score significantly lower on all well-being measures. Being a high-hours carer significantly reduces well-being for all measures, except consumption, once more suggesting that consumption may be picking up need. The presence of children is detrimental to income, consumption, CASP-19, domain satisfaction and life satisfaction (except 0-4 year olds) but for the GHQ measures only significantly so for 12 to 18 year olds. Having high contact with friends and family and neighbours is significantly positive for all the subjective measures, in contrast to the negative relationship with income and consumption noted above. In some important ways therefore, determinants of well-being differ when using income and consumption as outcome measures compared to subjective measures. In general determinants across different subjective measures are very similar. However, there are some interesting differences in the case of income, education, and gender. These are discussed below.

In the OLS models, household income has a positive relationship with all measures, and is significant for all except the GHQ ‘Caseness’. The large standard errors within the fixed effects models give large confidence intervals, as can be seen in Figure 10. Unfortunately, the explanatory power of fixed effects models is generally weak due to limited within person variability over what is a fairly short time period. Income remains significant for life satisfaction and domain-weighted satisfaction (and the coefficient changes only slightly), but not for the GHQ measures. The GHQ positive questions therefore appear to have a smaller relationship with income than other SWB measures. Looking again at the exact questions (Appendix 3.B), this is not surprising. GQH positive includes questions about concentration,
enjoying day-to-day activities and making decisions where as GHQ negative includes questions about loosing sleep due to worry, feeling under constant strain, not being able to overcome difficulties. The latter are more likely to identify those experiencing financial difficulties than the former.

**Figure 10: Coefficients for income on a range of SWB dependent variables**

Note: * = wave 11, other data refers to wave 9&14; error bars show 95% confidence intervals. For ease of display GHQ Likert and GHQ Caseness are not shown for wave 11, however, these are similar to wave 9&14.

The possibility of reverse causality creates a problem for interpreting coefficients from well-being regressions. In the case of income, for example, higher well-being may subsequently result in improved income through increased productivity. This effect may be stronger for some measures than others. For example, higher mental health or higher HRQoL may feed into productivity more strongly than positive life appraisal, however, this does not seem to be the case here. In the case of income, there is some evidence of a direction of causality running from higher well-being to higher income. For example, Graham *et al.* (2004) find that in Russia happiness which cannot be explained by socio-economic and demographic characteristics in 1995 predicted income in 2000. Diener *et al.* (2002) found that the level of cheerfulness of new college students was positively related to incomes 16 years later, even when controlling for
parents income. Marks and Fleming (1999) find some evidence that SWB in the previous period is positively related to income in The Australian Youth in Transition panel data. However, none of these studies control for personality or individual fixed effects, nor health in the case of Marks and Fleming, hence the role of SWB in determining future income is unclear.

Furthermore, whilst these do suggest some causality running from SWB to income, this is generally income some time in the future, and where the income effect arises from increased productivity we would expect a time delay before this impacted upon promotion and salary.

The relationship between education qualifications and well-being shows very interesting differences between the measures. The relationship is positive for income and consumption, the CASP-19, future SWB and past SWB, however, it is negative for all other measures in the OLS models, significantly so for life satisfaction, domain-weighted life satisfaction and at some education levels for the GHQ Caseness and Sf-6d. These differences are demonstrated in Figure 11 which shows the impact of having a degree.

**Figure 11: Coefficients for having a degree on a range of SWB dependent variables**

Note: * = wave 11, other data refers to wave 9&14; error bars show 95% confidence intervals.
Once more large standard errors in the fixed effects models result in lack of significance. This is not surprising since there is likely to be very little variation in qualifications over 5 years once respondents have reached adulthood. A within person comparison of the impact of further and higher education is, however, only part of the story, showing the impact of completing education, or returning to education in adulthood. Although not significant, the positive coefficients hint at a positive benefit arising from completing a degree. From a policy perspective, between person comparisons of education levels may be more relevant. The negative coefficients on GHQ and life satisfaction suggest those with a degree, all else equal, have poorer mental health and lower overall life satisfaction.

Figure 12 shows that when other factors are controlled for men score significantly lower on the CASP-19, life satisfaction and past and future SWB measures, but significantly higher for GHQ scores and the Sf-6d, confirming findings found in the literature discussed above.

**Figure 12: Coefficients for being male on a range of SWB dependent variables**

![Figure 12: Coefficients for being male on a range of SWB dependent variables](image-url)

Note: * = wave 11, other data refers to wave 9 & 14; error bars show 95% confidence intervals.
Coefficients on marriage status are reasonably similar across different SWB dependent variables. Divorce/separation generally has a strong negative effect across all measures. However, it only remains significant for domain satisfactions in the fixed effects models. The largest effect arises on past SWB, which again is closer to life satisfaction than is future SWB. The life satisfaction coefficient reduces in the fixed effects models, suggesting the possibility of some selection effects contributing to the impact of marriage status.

The coefficient sizes sometimes reduce in the fixed effects models, and significance is always reduced, suggesting greater between person variation than within person variation. Given the limited length of the panel this is not at all surprising.

Minor differences between measures are unlikely to have substantial policy relevance. However, in the case of gender and education, it is not only the magnitude of the effect that differs between different well-being measures but the direction of relationship. This raises the question of whether the different measures are measuring the same concept, or whether another explanation for these differences can be found.

As noted above the gender differences between well-being measures may arise if women either experience, or have greater willingness to report, emotional diversity and mental health problems. It is clearly important to be aware of these gender differences in the SWB measures, which appear to suggest that that psychological health measures are picking up a slightly different concept to the evaluative measures of well-being, or at the very least are subject to different reporting issues. However, as a fixed characteristic, gender is not the target of any government policy.

The education findings present more of a puzzle. We need to explain why the CASP-19, past SWB and future SWB, asked in wave 11, show a different relationship to overall life satisfaction, domain-weighted satisfaction, and the GHQ. A distinction based on evaluative measures versus psychological health measures cannot explain these findings.

Understanding the SWB-education relationship is of particular relevance for government policy. The UK government is committed to a target of 50% of all 18-30 years participating in higher education by 2010, an increase from 9% in 1999/2000 and 43% in 2005/2006 (Department for Education and Skills, 2007: 28). This involves considerable commitment of individual and public resources. The wisdom of such expansion may questioned if further education in its current form does not lead to higher well-being.

The negative finding on the GHQ may point to job-related stresses of employment following the acquisition of a degree or equivalent. The positive relationship with CASP-19 may arise
because it measures more functioning/flourishing aspects of well-being which may be more closely aligned to education. For example, one question asks if the respondent agrees with the statement “I choose to do things that I have never done before”. However, that would still not explain the positive relationship between education and past and future SWB.

There are good reasons why education may be beneficial to an individual’s SWB. Education is likely to increase an individual’s chances of having, among other things, higher income, increased wealth, improved employment opportunities and better health. It may improve decision-making skills, lead to increased confidence and improved conflict management. Ross and van Willigen (1997) propose that education enhances SWB due to access to supportive relationships, particularly marriage, and access to non-alienated paid work or “work that gives people the freedom from routinization, monotony, and external control on the one hand, and a chance to use their skills, develop as a person, and learn new things on the other” (Ross and van Willigen, 1997: 275).

Simultaneously, there may be reasons why education is detrimental to SWB. As discussed above, for any level of objective circumstances education may raise expectations therefore increasing the likelihood of the individual assessing those circumstances less favourably. Ross and van Willigen (1997) also found that controlling for non-alienated work and economic resources education increases job dissatisfaction. Psychological well-being and measures of distress may be less confounded by high expectations than measures of satisfaction, however, in this case the GHQ measure also shows a negative relationship. Nevertheless, most other studies showing a negative relationship between education and SWB have tended to use life satisfaction measures. There may be something about the word ‘satisfaction’ which restricts the highly educated from answering at the top of the scale, for example, being fully satisfied may be interpreted as a failure of aspiration or ambition.

The findings for Britain, in terms of life satisfaction and mental health, present a less positive picture of education than in other countries. However, it should be noted that within the BHPS education has a positive effect when health, job status, marriage status and income are not controlled for. Education remains significant (in random effects) when income is controlled for, but not once self-rated health is introduced.

Another novel explanation is to consider the number of response options within the scales used for the different measures. The satisfaction questions are scored on a 1-7 scale, whereas the SWB past and future scales are scored by the aggregate of two 0-3 scales. Although a direct comparison is impossible because life satisfaction and the CASP-19 questions are never asked in the same wave, the average life satisfaction scores (mean of 70% of the scale) are lower than
past and future SWB (mean of 78% and 80% of the scale). This would arise if people were more willing to use the top of the scale in a 0-3 scale than a 1-7 scale.

As briefly discussed above, the evidence on education and SWB appears in many ways to be contradictory, a number of studies finding a positive affect, many finding no significant effect and some finding a negative effect. To help address the education puzzle, the literature is reconsidered to see if positive relationships are mainly found when few option response scales are used, and negative relationships when more numerous response scales are used. Studies which find a positive impact (Table 15), negative impact (Table 16) and non-significant impact (Table 17) on SWB are shown in Appendix 3.F, along with studies showing the impact of education on mental health (Table 18).

The majority of studies finding a positive effect of education use scales with 5 points or less, whereas the majority of studies finding a negative or insignificant impact of education use scales with 7 points or more, although this relationship is far from clear cut. These findings may be explained if education reduces the chances of having very high life satisfaction but simultaneously reduces the chances of having low satisfaction. On a 3 point scale responses in the top category may reasonably be assumed to be given if true life satisfaction is somewhere in the top third of an individual’s assessment of the best and worst life possible. It may be that the more educated individuals fall at the bottom end of this top third but this cannot be revealed by the scale. With more response options it would be possible to see whether education reduces the probability of reporting very high levels of life satisfaction. This can be explored by considering the factors that influence the probability of getting a 7/7 on the life satisfaction scale, compared to other scores.

Whether an individual scores 7/7 is a binary variable that can be modelled in terms of an underlying latent variable (or tendency to have the highest level of life satisfaction).

\[ LS7_{it}^* = \beta Z_{it} + \alpha_t + \varepsilon_{it} \]  

(7)

where we observe \( LS7_{it} = 1 \) if \( LS7_{it}^* > 0 \) and \( LS7_{it} = 0 \) otherwise.

As noted above, Chamberlain (1980) has shown that it is possible to use Maximum Likelihood estimation, control for fixed effects and get consistent estimates for coefficients in models with dichotomous dependent variables.

The remainder of the life satisfaction scale (1-6) can be treated cardinally, thereby controlling for fixed effects, using OLS fixed effects, as in (2) above. Alternatively it may be treated as
ordinal using a random effects ordered probit model (see Maddala, 1983). The latter also assumes life satisfaction is a latent unobserved variable, LS*.

\[ LS_{it}^* = \beta Z_{it} + \alpha_i + \epsilon_{it} \]  

(7)

where we observe:

\[
\begin{align*}
LS_{it} &= 1 \text{ if } -\infty \leq LS^* < \mu_1 \\
&= 2 \text{ if } \mu_1 < LS^* < \mu_2 \\
&= 3 \text{ if } \mu_2 < LS^* < \mu_3 \text{ etc.}
\end{align*}
\]

The probability of any particular individual being in a particular category (e.g. life satisfaction score of 5) is calculated as the probability of a linear function of the independent variables being within the cut points or thresholds (\(\mu_1, \mu_2, \mu_3\) etc.) which are also estimated within the model.

Using a probit model with individual random effects also requires the assumption that the individual effects are not correlated with the independent variables (i.e. \(E(\alpha_i Z_{it}) = 0\)).

To explore the education effects in more detail, highest education qualification is further disaggregated. To maximise sample size, gross household current income is used rather than net, allowing wave 15 to be included. Controlling for the unobserved individual effect is important when considering the impact of education, since education achievement is likely to be correlated to personality factors. However, the use of fixed effects limits the analysis to within person variance, or those who have changed education during the sample years. Table 4 shows that, as anticipated, in most cases, education qualifications compared to no-qualifications leads to a positive but non-significant effect. The outlier is nursing, which results in negative, and in some cases significant effect. However, these findings do not help solve the puzzle of negative impact of education for life satisfaction but not past and future SWB which arose when fixed effects were not controlled for.

<table>
<thead>
<tr>
<th>Compared to no qualifications</th>
<th>Conditional logit coefficient, FE (7/7) (se)</th>
<th>Marginal effects (7/7)</th>
<th>OLS, FE (1-6 only) (se)</th>
<th>OLS, FE (1-7) (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td>.8429(.5314)</td>
<td>.2067</td>
<td>.0552(.1013)</td>
<td>.1575(.0953)</td>
</tr>
<tr>
<td>First degree</td>
<td>.1539(.3193)</td>
<td>.0381</td>
<td>.0551(.0709)</td>
<td>.0990(.0633)</td>
</tr>
<tr>
<td>Teaching</td>
<td>.6604(1.0584)</td>
<td>.1635</td>
<td>-.0622(.2290)</td>
<td>.0134(.2112)</td>
</tr>
<tr>
<td>Other higher</td>
<td>.3609(2.2000)</td>
<td>.0894</td>
<td>.0036(.0574)</td>
<td>.0567(.0491)</td>
</tr>
<tr>
<td>Nursing</td>
<td>.4535(.5556)</td>
<td>.1128</td>
<td>-.2342(.1178)*</td>
<td>-.1240(.1101)</td>
</tr>
<tr>
<td>A level</td>
<td>.4576(2.461)</td>
<td>.1137</td>
<td>.0038(.0609)</td>
<td>.0573(.0531)</td>
</tr>
<tr>
<td>O level</td>
<td>.3718(2.384)</td>
<td>.0922</td>
<td>-.0117(.0604)</td>
<td>.0214(.0525)</td>
</tr>
<tr>
<td>Commercial</td>
<td>.2847(.4932)</td>
<td>.0707</td>
<td>-.0385(.1207)</td>
<td>.0068(.1065)</td>
</tr>
<tr>
<td>CSE</td>
<td>.5781(.3387)</td>
<td>.1435</td>
<td>-.1626(.0901)</td>
<td>-.0620(.0776)</td>
</tr>
<tr>
<td>Apprentice</td>
<td>.4194(.6017)</td>
<td>.1043</td>
<td>.0277(.1630)</td>
<td>.1394(.1424)</td>
</tr>
</tbody>
</table>
Results which incorporate both the within person and between person variation more clearly highlight the source of the negative education coefficient on life satisfaction (Table 5). The random effects probit model shows negative education coefficients for the probability of a life satisfaction response of 7, compared to all other responses. The negative effect of education increases with education level, and is particularly strong for degree level education, which has a marginal effect of -0.0336, suggesting a move from no education to degree level reduces the probability of reporting 7 out of 7 by 3 percent when all variables are taken at their mean value.

However, when taking only individuals who have only ever responded between 1 and 6 during their time in the panel, education shows a positive relationship to life satisfaction. When an ordered probit is run on the whole sample education once again has a negative coefficient, suggesting the negative impact of higher education on attaining a 7/7 is sufficiently strong to result in an overall negative relationship.

This pattern is also apparent when treating the life satisfaction scale as continuous through the use of GLS random effects models, shown in Table 6. More education reduces the chance of responding at the top of the life satisfaction scale, yet simultaneously reduces the chances of answering at the low end of the scale. This same pattern is found for the effect of education across all domains.

### Table 5: Coefficients for education level from RE models on life satisfaction

<table>
<thead>
<tr>
<th>Compared to no qualifications</th>
<th>Probit coefficient, re (7/7) (se)</th>
<th>Marginal effects (7/7)</th>
<th>Ordered probit coefficients, re (never responded 7) (se)</th>
<th>Ordered probit coefficients re (Full sample) (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td>-1.1024(.1151)**</td>
<td>-0.270**</td>
<td>.3446(.0595)**</td>
<td>-.2317(.0541)**</td>
</tr>
<tr>
<td>First degree</td>
<td>-1.1744(.0622)**</td>
<td>-0.336**</td>
<td>.3411(.0380)**</td>
<td>-.2175(.0325)**</td>
</tr>
<tr>
<td>Teaching</td>
<td>-.8011(.0926)**</td>
<td>-0.246**</td>
<td>.3179(.0668)**</td>
<td>-.1604(.0577)**</td>
</tr>
<tr>
<td>Other higher</td>
<td>-.6980(.0432)**</td>
<td>-.0312**</td>
<td>.2583(.0317)**</td>
<td>-.1461(.0260)**</td>
</tr>
<tr>
<td>Nursing</td>
<td>-.5915(.1150)**</td>
<td>-0.0212**</td>
<td>.0908(.0794)**</td>
<td>-.2295(.0685)**</td>
</tr>
<tr>
<td>A level</td>
<td>-.7342(.0503)**</td>
<td>-0.0277**</td>
<td>.2886(.0352)**</td>
<td>-.1558(.0291)**</td>
</tr>
<tr>
<td>O level</td>
<td>-.5403(.0438)**</td>
<td>-0.0245**</td>
<td>.2244(.0331)**</td>
<td>-.1451(.0272)**</td>
</tr>
<tr>
<td>Commercial</td>
<td>-.3469(.0805)**</td>
<td>-0.0156**</td>
<td>.1033(.0720)**</td>
<td>-.0987(.0544)**</td>
</tr>
<tr>
<td>CSE</td>
<td>-.2327(.0756)**</td>
<td>-0.0116**</td>
<td>.1224(.0554)**</td>
<td>-.0858(.0468)**</td>
</tr>
<tr>
<td>Apprentice</td>
<td>-.2409(.0880)**</td>
<td>-0.0122**</td>
<td>.1940(.0796)**</td>
<td>-.0160(.0675)**</td>
</tr>
<tr>
<td>Other education</td>
<td>-.3394(.1768)</td>
<td>-0.0151**</td>
<td>.1396(.1111)</td>
<td>-.0236(.0990)**</td>
</tr>
<tr>
<td>Observations</td>
<td>88247</td>
<td></td>
<td>57864</td>
<td>88247</td>
</tr>
<tr>
<td>Individuals</td>
<td>17958</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-26332.22</td>
<td></td>
<td>-70096.342</td>
<td>-119532.86</td>
</tr>
</tbody>
</table>
Note: ** p < 0.01, * p < 0.05, control variables as above, sample excludes NI and those still at school.

It would be possible to further explore this relationship through consideration of each step on the life satisfaction scale separately, e.g. response of 1 compared to all scores above 1, a response of 1 or 2 compared to all other scores above 2 etc. These probabilities can be estimated simultaneously using a random effects generalized ordered probit model (Boes, 2006; Boes and Winkelmann, 2006). This has the advantage of more accurate standard errors. The generalized model, estimated using maximum likelihood, relaxes the parallel regression assumption of the random effects ordered probit models and allows the effects of education on SWB to vary in different parts of the life satisfaction distribution. However, these results still rest on the appropriateness of the assumption that the individual effects are not correlated with the independent variables.

Table 6: Education coefficients from a generalised ordered probit RE model on life satisfaction

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td>1.180</td>
<td>0.174</td>
<td>0.002</td>
<td>0.131</td>
<td>-0.122</td>
<td>-0.960</td>
</tr>
<tr>
<td></td>
<td>(.266)**</td>
<td>(.099)</td>
<td>(.075)</td>
<td>(.066)*</td>
<td>(.061)*</td>
<td>(.081)**</td>
</tr>
<tr>
<td>First degree</td>
<td>0.459</td>
<td>0.139</td>
<td>0.054</td>
<td>0.112</td>
<td>-0.120</td>
<td>-0.979</td>
</tr>
<tr>
<td></td>
<td>(.085)**</td>
<td>(.056)*</td>
<td>(.044)</td>
<td>(.038)**</td>
<td>(.036)**</td>
<td>(.046)**</td>
</tr>
<tr>
<td>Teaching</td>
<td>0.413</td>
<td>0.174</td>
<td>0.148</td>
<td>0.147</td>
<td>-0.038</td>
<td>-0.721</td>
</tr>
<tr>
<td></td>
<td>(.146)**</td>
<td>(.098)</td>
<td>(.077)</td>
<td>(.066)*</td>
<td>(.062)</td>
<td>(.071)**</td>
</tr>
<tr>
<td>Other higher</td>
<td>0.263</td>
<td>0.233</td>
<td>0.115</td>
<td>0.082</td>
<td>-0.103</td>
<td>-0.581</td>
</tr>
<tr>
<td></td>
<td>(.056)**</td>
<td>(.044)**</td>
<td>(.035)**</td>
<td>(.030)**</td>
<td>(.029)**</td>
<td>(.033)**</td>
</tr>
<tr>
<td>Nursing</td>
<td>0.183</td>
<td>0.099</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.258</td>
<td>-0.559**</td>
</tr>
<tr>
<td></td>
<td>(.145)</td>
<td>(.113)</td>
<td>(.093)</td>
<td>(.081)</td>
<td>(.076)**</td>
<td>(.086)</td>
</tr>
<tr>
<td>A level</td>
<td>0.343</td>
<td>0.239</td>
<td>0.191</td>
<td>0.110</td>
<td>-0.136</td>
<td>-0.599</td>
</tr>
<tr>
<td></td>
<td>(.068)**</td>
<td>(.052)**</td>
<td>(.040)**</td>
<td>(.034)**</td>
<td>(.032)**</td>
<td>(.038)**</td>
</tr>
<tr>
<td>O level</td>
<td>0.283</td>
<td>0.169</td>
<td>0.105</td>
<td>0.025</td>
<td>-0.115</td>
<td>-0.441</td>
</tr>
<tr>
<td></td>
<td>(.057)**</td>
<td>(.045)**</td>
<td>(.036)**</td>
<td>(.031)</td>
<td>(.030)**</td>
<td>(.033)**</td>
</tr>
<tr>
<td>Commercial</td>
<td>0.159</td>
<td>0.145</td>
<td>0.113</td>
<td>0.023</td>
<td>-0.102</td>
<td>-0.309</td>
</tr>
<tr>
<td></td>
<td>(.118)</td>
<td>(.093)</td>
<td>(.073)</td>
<td>(.063)</td>
<td>(.060)</td>
<td>(.065)**</td>
</tr>
<tr>
<td>CSE</td>
<td>0.265</td>
<td>0.193</td>
<td>0.114</td>
<td>0.013</td>
<td>-0.086</td>
<td>-0.250</td>
</tr>
<tr>
<td></td>
<td>(.106)*</td>
<td>(.081)*</td>
<td>(.063)</td>
<td>(.054)</td>
<td>(.052)</td>
<td>(.059)**</td>
</tr>
<tr>
<td>Apprentice</td>
<td>0.179</td>
<td>0.189</td>
<td>0.072</td>
<td>0.094</td>
<td>-0.011</td>
<td>-0.172</td>
</tr>
<tr>
<td></td>
<td>(.123)</td>
<td>(.105)</td>
<td>(.086)</td>
<td>(.075)</td>
<td>(.071)</td>
<td>(.075)**</td>
</tr>
<tr>
<td>Other education</td>
<td>0.151</td>
<td>0.186</td>
<td>-0.017</td>
<td>0.108</td>
<td>0.034</td>
<td>-0.108</td>
</tr>
<tr>
<td></td>
<td>(.189)</td>
<td>(.168)</td>
<td>(.133)</td>
<td>(.120)</td>
<td>(.113)</td>
<td>(.121)</td>
</tr>
</tbody>
</table>

Note: ** p < 0.01, control variables as above, sample excludes NI and those still at school. Observations: 88,247

42 Generalised ordered logit models are usually applied to situations where it is necessary to account for measurement error arising from state-dependent reporting bias (the use of varying cut points) (Lindeboom and Doorslaer, 2004). Situations where the effect of x on y varies across the range of y can often be dealt with through appropriate transformation of the y variable. However, this is not suitable for ordinal models, and problematic here since it is not theorised that other variables necessarily share a similar relationship where there effect varies according to the level of y.
The generalised ordered probit model confirms the negative impact of education on responses at the top of the scale. The pattern can be clearly seen in the impact of having A levels, which improves the chances of reporting higher up the life satisfaction scale when evaluated from the bottom of the scale, but reduces the chances of being higher up the life satisfaction scale when evaluated from the high end of the scale. This finding could go some way to explaining the apparent inconsistencies within the relationships between education and well-being measures found both in the literature and the BHPS analysis here. However, it remains unclear whether this effect arises due to the concept of ‘satisfaction’ or the use of a 1 to 7 scale. More work on the differences in expectations across different education groups would help unravel this further. It would also be interesting to know whether these patterns were also picked up in biomarkers, such as levels of cortisol. The possibility that characteristics may impact differently at different points in a well-being scale underlies the importance of having similar response scales for different SWB questions in order to distinguish between conceptual differences between well-being measures and those caused by differences in response scales.

3.7 Summary

This chapter has considered empirically the degree of similarity between different well-being measures. Income and consumption are shown to diverge from other SWB measures. Bivariate analysis shows that for some groups income and consumption present a very different picture than the subjective measures. Many of the groups that show considerable variation are politically important groups (elderly, single parents, disabled, for example) for whom policy makers are likely to need an accurate, valid measure of well-being.

Multivariate analysis finds that for many variables found in the literature to impact upon SWB, including unemployment, health status, being a high-load carer, the extent of social contact, widowhood, and divorce, the different SWB measures offer a consistent message. However, some differences are still apparent across gender and education. This may imply they are measuring slightly different concepts. Unfortunately, neither gender nor education are well suited to fixed effects models, making it unclear what role personality is playing in these findings. Analysis has shown that for education part of the variation across different measures arises because of the reduced probability of those with higher education answering at the top end of the 1-7 satisfaction scales.

Some measures of well-being are more sensitive at the low well-being or unhappiness part of the scale (such as the GHQ ‘Caseness’), others are more sensitive at the high well-being or happiness part of the scale. Some determinants of well-being may operate at both ends of the spectrum e.g. unemployment or divorce reduces happiness and increases unhappiness. Other
determinants may impact only at one end e.g. income may reduce unhappiness but not increase happiness (Boes and Winkelmann, 2006). This is an area that warrants further investigation.

This chapter has demonstrated that an individual’s well-being is dependent upon the choice of well-being measure. This is especially so for some important, potentially vulnerable, groups. Income and consumption are out of step with a range of other subjective measures which brings into doubt their validity as measures of individual well-being. Although broadly speaking the SWB measures tell a similar story, there are some important variations. A greater understanding of these inconsistencies would help move the SWB measures forward as prospective measures for public policy.
Table 7: Correlations between income and consumption, wave 14

<table>
<thead>
<tr>
<th>Income measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HH, annual, g.</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. HH, eq (M) annual, g.</td>
<td>0.87</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. HH, eq (O) annual, g.</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. IND, annual, g.</td>
<td>0.62</td>
<td>0.72</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. HH, current, g.</td>
<td>0.87</td>
<td>0.73</td>
<td>0.75</td>
<td>0.55</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. HH, eq (M), current, g.</td>
<td>0.75</td>
<td>0.84</td>
<td>0.84</td>
<td>0.64</td>
<td>0.88</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. HH, eq (O), current, g.</td>
<td>0.77</td>
<td>0.84</td>
<td>0.84</td>
<td>0.63</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. IND, current, g.</td>
<td>0.54</td>
<td>0.60</td>
<td>0.60</td>
<td>0.86</td>
<td>0.66</td>
<td>0.75</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. HH, annual, n.</td>
<td>0.98</td>
<td>0.83</td>
<td>0.87</td>
<td>0.59</td>
<td>0.86</td>
<td>0.72</td>
<td>0.75</td>
<td>0.51</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. HH, eq (M) annual, n.</td>
<td>0.83</td>
<td>0.98</td>
<td>0.97</td>
<td>0.69</td>
<td>0.70</td>
<td>0.83</td>
<td>0.82</td>
<td>0.58</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. HH, eq (O), annual, n.</td>
<td>0.87</td>
<td>0.97</td>
<td>0.98</td>
<td>0.68</td>
<td>0.73</td>
<td>0.83</td>
<td>0.83</td>
<td>0.57</td>
<td>0.87</td>
<td>0.99</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. HH, current, net</td>
<td>0.87</td>
<td>0.71</td>
<td>0.74</td>
<td>0.53</td>
<td>0.98</td>
<td>0.84</td>
<td>0.87</td>
<td>0.62</td>
<td>0.89</td>
<td>0.71</td>
<td>0.74</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. HH, eq (M) current, n.</td>
<td>0.74</td>
<td>0.84</td>
<td>0.84</td>
<td>0.62</td>
<td>0.85</td>
<td>0.97</td>
<td>0.97</td>
<td>0.71</td>
<td>0.74</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. HH, eq (O), current n.</td>
<td>0.77</td>
<td>0.84</td>
<td>0.84</td>
<td>0.62</td>
<td>0.87</td>
<td>0.97</td>
<td>0.97</td>
<td>0.71</td>
<td>0.77</td>
<td>0.85</td>
<td>0.86</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. HH, consumption</td>
<td>0.63</td>
<td>0.44</td>
<td>0.46</td>
<td>0.32</td>
<td>0.61</td>
<td>0.44</td>
<td>0.46</td>
<td>0.30</td>
<td>0.61</td>
<td>0.39</td>
<td>0.42</td>
<td>0.61</td>
<td>0.41</td>
<td>0.43</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. HH, eq (M) consumption</td>
<td>0.43</td>
<td>0.50</td>
<td>0.49</td>
<td>0.36</td>
<td>0.41</td>
<td>0.49</td>
<td>0.48</td>
<td>0.33</td>
<td>0.39</td>
<td>0.47</td>
<td>0.46</td>
<td>0.39</td>
<td>0.47</td>
<td>0.46</td>
<td>0.79</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>17. HH, eq. (O) consumption</td>
<td>0.47</td>
<td>0.51</td>
<td>0.51</td>
<td>0.36</td>
<td>0.45</td>
<td>0.50</td>
<td>0.50</td>
<td>0.34</td>
<td>0.44</td>
<td>0.48</td>
<td>0.48</td>
<td>0.43</td>
<td>0.48</td>
<td>0.48</td>
<td>0.84</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>18. IND, consumption</td>
<td>0.35</td>
<td>0.36</td>
<td>0.36</td>
<td>0.38</td>
<td>0.34</td>
<td>0.35</td>
<td>0.36</td>
<td>0.36</td>
<td>0.34</td>
<td>0.34</td>
<td>0.35</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.46</td>
<td>0.46</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Note: Where IND is individual, HH is household, eq (M) is equivalised using McClements, eq (O) is using the modified OECD scale, g is gross and n is net. Observations = 9,517, those with complete income and consumption data.
### Table 8: Correlations between well-being measures, wave 14

<table>
<thead>
<tr>
<th></th>
<th>Income</th>
<th>Consumption</th>
<th>GHQ ‘Likert’</th>
<th>GHQ ‘Caseness’</th>
<th>GHQ Positive</th>
<th>GHQ Negative</th>
<th>Life Satisfaction</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumption</td>
<td>0.501</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ ‘Likert’</td>
<td>0.088</td>
<td>0.059</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ ‘Caseness’</td>
<td>0.071</td>
<td>0.036</td>
<td>0.896</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ Positive</td>
<td>0.087</td>
<td>0.084</td>
<td>0.864</td>
<td>0.796</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ Negative</td>
<td>0.075</td>
<td>0.033</td>
<td>0.941</td>
<td>0.829</td>
<td>0.643</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>0.076</td>
<td>0.040</td>
<td>0.554</td>
<td>0.499</td>
<td>0.435</td>
<td>0.551</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Domains</td>
<td>0.117</td>
<td>0.068</td>
<td>0.517</td>
<td>0.463</td>
<td>0.387</td>
<td>0.527</td>
<td>0.766</td>
<td>1.000</td>
</tr>
<tr>
<td>Sf-6d</td>
<td>0.184</td>
<td>0.204</td>
<td>0.516</td>
<td>0.491</td>
<td>0.460</td>
<td>0.477</td>
<td>0.382</td>
<td>0.431</td>
</tr>
</tbody>
</table>

Note: Observations = 9,022, those with complete data.

### Table 9: Correlations between well-being measures, wave 11

<table>
<thead>
<tr>
<th></th>
<th>Income</th>
<th>Consumption</th>
<th>GHQ ‘Likert’</th>
<th>GHQ ‘Caseness’</th>
<th>GHQ Positive</th>
<th>GHQ Negative</th>
<th>CASP-19</th>
<th>SWB Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumption</td>
<td>0.524</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ ‘Likert’</td>
<td>0.075</td>
<td>0.052</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ ‘Caseness’</td>
<td>0.051</td>
<td>0.017</td>
<td>0.902</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ Positive</td>
<td>0.077</td>
<td>0.083</td>
<td>0.869</td>
<td>0.800</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ Negative</td>
<td>0.062</td>
<td>0.023</td>
<td>0.941</td>
<td>0.838</td>
<td>0.650</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASP-19</td>
<td>0.201</td>
<td>0.190</td>
<td>0.577</td>
<td>0.490</td>
<td>0.481</td>
<td>0.556</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>SWB past</td>
<td>0.115</td>
<td>0.086</td>
<td>0.425</td>
<td>0.373</td>
<td>0.320</td>
<td>0.433</td>
<td>0.721</td>
<td>1.000</td>
</tr>
<tr>
<td>SWB future</td>
<td>0.134</td>
<td>0.134</td>
<td>0.512</td>
<td>0.436</td>
<td>0.439</td>
<td>0.485</td>
<td>0.793</td>
<td>0.646</td>
</tr>
</tbody>
</table>

Note: Observations = 9,930.
APPENDIX 3.B

*General Health Questionnaire 12 (GHQ12) Scale*

"We would like to know how your health has been in general, over the past few weeks. Please answer the following questions by circling the number that best applies to you. Have you recently…"

- Been able to concentrate on whatever you are doing?
  - (Better/same/Less/Much less) than usual.
- Lost much sleep over worry? *
- Felt constantly under strain? *
- Felt that you couldn't overcome your difficulties? *
- Been feeling unhappy and depressed? *
- Been losing self-confidence in yourself? *
- Been thinking of yourself as a worthless person? *
  - (Not at all/No more/Rather more/Much more) than usual.
- Felt that you were playing a useful part in things?
- Felt capable of making decisions about things?
- Been able to enjoy your normal day-to-day activities?
- Been able to face up to your problems?
- Been feeling reasonably happy, all things considered?
  - (More so/Same as/Less/Much less) than usual.

Items marked * are reverse coded. Full scale (Likert) scores between 0 and 36. An alternative scoring system allocates 1 if the symptom is present (responses less or much less for positives, and rather more or much for the negatives) and then aggregates the number of symptoms identified. This is known as the ‘Caseness’ score, which refers to the use of cut off points to divide the population into ‘cases’ and ‘normals’, where ‘Caseness’ expresses a high probability that the respondent will be found to have a psychiatric illness at further investigation. Within clinical studies the cut off point is chosen such that it best discriminates between two groups. For the GHQ12 the threshold for identification of cases usually varies between three and six (Goldberg and Williams, 1988).
**CASP-19 Scale**

“Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, do you think this applies to you:”

0 Often; 1 Sometimes; 2 Not often; 3 Never

**Control**

My age prevents me from doing the things I would like to.*
I feel that what happens to me is out my control.*
I feel free to plan for the future.
I feel left out of things.*

**Autonomy**

I can do the things I want to do.
Family responsibilities prevent me from doing things I want to do.*
I feel that I can please myself what I can do.
My health stops me from doing the things I want to do.*
Shortage of money stops me from doing the things I want to do.*

**Pleasure**

I look forward to each day.
I feel that my life has meaning.
I enjoy the things that I do.
I enjoy being in the company of others.
On balance, I look back on my life with a sense of happiness.

**Self-realization**

I feel full of energy these days.
I choose to do things that I have never done before.
I feel satisfied with the way my life has turned out.
I feel that life is full of opportunities.
I feel that the future looks good for me.

Items marked * are reverse coded.
Sum of all items (0 to 57)
Figure 13: Histograms of well-being measures
### Table 10: Description of variables, waves 7-14, unless stated

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
<th>Mean (sd) or % of cases</th>
<th>Min</th>
<th>Max</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household income, net current</td>
<td>Weekly household net income. Taken from variable ‘hhneti’. This is deflated to 1987 prices using RPI.</td>
<td>£254.34 (182.33)</td>
<td>-£16.45</td>
<td>£5,993.86</td>
<td>85,467</td>
</tr>
<tr>
<td>Household income, gross, current</td>
<td>Household gross income, the month before the interview. Taken from variable ‘fihhmn’. This is deflated to 1987 prices using the RPI. Includes imputed data.</td>
<td>£1,442.38 (1143.51)</td>
<td>0</td>
<td>£4,476.24</td>
<td>96,343</td>
</tr>
<tr>
<td>Mean regional household income, net current</td>
<td>Weekly household net income, average for the region, excluding the individual household. Taken from ‘hhneti’. Deflated to 1987 priced using RPI.</td>
<td>£254.34 (31.37)</td>
<td>£187.86</td>
<td>361.96</td>
<td>85,467</td>
</tr>
<tr>
<td>Ratio of own household to regional household income</td>
<td>Ratio of the log of net current household income over the log of average regional household income (net current) for the region (excluding current individual)</td>
<td>96.25 (12.86)</td>
<td>-53.49</td>
<td>155.22</td>
<td>84,954</td>
</tr>
<tr>
<td>Mean regional household income, net current by education and age group</td>
<td>Average regional income (net, current) by gender, age (4 groups) and education (7 groups)</td>
<td>£254.42 (83.25)</td>
<td>-£15.33</td>
<td>£3,250.62</td>
<td>85,319</td>
</tr>
<tr>
<td>Mean regional consumption</td>
<td>Average consumption (total of variables available) for the region (excluding individual)</td>
<td>£5,019.57 (675.45)</td>
<td>£3405.62</td>
<td>£7,175.48</td>
<td>63,926</td>
</tr>
<tr>
<td><strong>Financial situation</strong></td>
<td>Taken from variable ‘fisit’, which asks “How well would you say you yourself are managing financially these days? Would you say you are...living comfortably/doing alright/just about getting by/finding it quite difficult/finding it very difficult”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living comfortably</td>
<td>Dummy 1 if response is living comfortably, 0 otherwise</td>
<td>31.5%</td>
<td>0</td>
<td>1</td>
<td>96,190</td>
</tr>
<tr>
<td>Doing alright</td>
<td>Dummy 1 if response is doing alright, 0 otherwise</td>
<td>37.3%</td>
<td>0</td>
<td>1</td>
<td>96,190</td>
</tr>
<tr>
<td>Just about getting by</td>
<td>Dummy 1 if response is just about getting by, 0 otherwise</td>
<td>24.0%</td>
<td>0</td>
<td>1</td>
<td>96,190</td>
</tr>
<tr>
<td>Finding it quite difficult</td>
<td>Dummy 1 if response is finding it quite difficult, 0 otherwise</td>
<td>5.0%</td>
<td>0</td>
<td>1</td>
<td>96,190</td>
</tr>
<tr>
<td>Finding it very difficult</td>
<td>Dummy 1 if response is finding it very difficult, 0 otherwise</td>
<td>2.1%</td>
<td>0</td>
<td>1</td>
<td>96,190</td>
</tr>
<tr>
<td><strong>Savings and debt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly savings</td>
<td>Monthly savings. Taken from ‘saved’, “About how much on average do you personally manage to save a month?” Deflated using RPI.</td>
<td>61.45 (187.17)</td>
<td>0</td>
<td>9999</td>
<td>91,363</td>
</tr>
<tr>
<td>Debt problem</td>
<td>Dummy 1 if repayment on loan or hp is heavy or somewhat of a burden on household, 0 otherwise. Taken from ‘xphpdf’.</td>
<td>11.60%</td>
<td>0</td>
<td>1</td>
<td>95,870</td>
</tr>
<tr>
<td><strong>Financial outlook</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pessimistic</td>
<td>Dummy 1 if financial outlook is worse than now.</td>
<td>8.85%</td>
<td>0</td>
<td>1</td>
<td>92,581</td>
</tr>
<tr>
<td>Optimistic</td>
<td>Dummy 1 if financial outlook is better than now.</td>
<td>31.15%</td>
<td>0</td>
<td>1</td>
<td>92,581</td>
</tr>
<tr>
<td><strong>Importance of money</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative importance money</td>
<td>Importance of income (‘lfimpb’), as a ratio to the average importance of good partnership, good friends and good health (‘lfimpg’, ‘lfimph’ and ‘lfimpa’). For waves 8 &amp; 13.</td>
<td>0.698</td>
<td>0</td>
<td>5</td>
<td>21,499</td>
</tr>
<tr>
<td><strong>Income inequality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional gini</td>
<td>Gini index calculated for each region and wave, based on current net household income. Defined as the ratio of the area between a Lorenz curve (cumulative distribution of income and cumulative distribution of the population) and the equality diagonal and the area below the equality diagonal.</td>
<td>0.339 (0.022)</td>
<td>0.286</td>
<td>0.488</td>
<td>95,343</td>
</tr>
<tr>
<td>Regional theil</td>
<td>Theil index is a generalised entropy measure with sensitivity parameter 1. Using individual data, this is determined by proportional distance from the mean.</td>
<td>0.193 (0.040)</td>
<td>0.132</td>
<td>0.589</td>
<td>95,343</td>
</tr>
</tbody>
</table>
Chapter 3  
Empirical differences between measures

<table>
<thead>
<tr>
<th>Regional co-efficient of variation</th>
<th>Co-efficient of variation calculated for each region and wave, based on current net household income. Standard deviation of the distribution divided by the mean.</th>
<th>0.689 (0.141)</th>
<th>0.505</th>
<th>1.864</th>
<th>95,343</th>
</tr>
</thead>
</table>

**Household composition**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Co-efficient</th>
<th>Degree</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. children 0-4</td>
<td>Number of children between ages 0 and 4 living in the household ('nch02' plus 'nch34')</td>
<td>0.136 (.402)</td>
<td>0</td>
<td>3</td>
<td>96,343</td>
</tr>
<tr>
<td>No. children 5-11</td>
<td>Number of children between ages 5 and 11 living in the household ('nch511')</td>
<td>0.248 (.589)</td>
<td>0</td>
<td>6</td>
<td>96,343</td>
</tr>
<tr>
<td>No. children 12-18</td>
<td>Number of children between ages 12 and 18 living in the household ('nch1215' plus 'nch1618')</td>
<td>0.220 (.539)</td>
<td>0</td>
<td>4</td>
<td>96,343</td>
</tr>
<tr>
<td>Adults</td>
<td>Household size ('hhsize') less number of children</td>
<td>2.22 (0.909)</td>
<td>0</td>
<td>9</td>
<td>96,343</td>
</tr>
</tbody>
</table>

**Education**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Co-efficient</th>
<th>Degree</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No qualifications</td>
<td>Dummy 1 if no education qualifications, 0 otherwise</td>
<td>22.24%</td>
<td>0</td>
<td>1</td>
<td>94,910</td>
</tr>
<tr>
<td>O level or commercial</td>
<td>Dummy 1 if highest education qualification is O level or commercial education or CSE grade 2-5, Apprenticeship or other, 0 otherwise.</td>
<td>26.92%</td>
<td>0</td>
<td>1</td>
<td>94,910</td>
</tr>
<tr>
<td>A level or similar</td>
<td>Dummy 1 if highest education qualification is A level or similar, 0 otherwise.</td>
<td>12.31%</td>
<td>0</td>
<td>1</td>
<td>94,910</td>
</tr>
<tr>
<td>Degree or equivalent</td>
<td>Dummy 1 if highest education qualification is higher degree, first degree, teaching qualification, other higher education qualification, or nursing qualification, 0 otherwise.</td>
<td>37.97%</td>
<td>0</td>
<td>1</td>
<td>94,910</td>
</tr>
<tr>
<td>Still at school</td>
<td>Dummy 1 if still at school with no qualifications, 0 otherwise</td>
<td>0.56%</td>
<td>0</td>
<td>1</td>
<td>94,910</td>
</tr>
<tr>
<td>High load carer</td>
<td>Dummy 1 if giving care for over 50 hours per week. Taken from ‘aidhrs’, hours of care both inside and outside the home (this does not include those respondents who rely that the care load varies but is over 20 hours).</td>
<td>1.8%</td>
<td>0</td>
<td>1</td>
<td>94,541</td>
</tr>
</tbody>
</table>

**Health**

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
<th>Co-efficient</th>
<th>Degree</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-rated health</td>
<td>Self-rated health taken from ‘hlstat’ (waves 1-8 and 10-15), which is coded 1-5 using responses ‘very poor/poor/fair/good/excellent’, and variable ‘hlsf1’ in wave 9 which is coded 1-5 using responses ‘poor/fair/good/very good/excellent’. These variables are recoded into 4 categories, ‘hlstat’ is recoded such that the first two categories are combined, and ‘hlsf1’ is recoded such that 3 and 4 are combined. This adopts a method employed by</td>
<td>0.689 (0.141)</td>
<td>0.505</td>
<td>1.864</td>
</tr>
</tbody>
</table>
### Empirical differences between measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
<th>Hernandez Quevedo et al. (2004)</th>
<th>Problems walking</th>
<th>Hospital visits</th>
<th>High GP visits</th>
<th>Disabled</th>
<th>Self-rated health poor or very poor</th>
<th>Self-rated health fair</th>
<th>Self-rated health good</th>
<th>Self-rated health excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dummy 1 if respondent not being able to walk 10 minutes (waves 6-8 and 10-15) or 1/2 mile (wave 9), 0 otherwise. Taken from ‘hlld’ and ‘hlsf3h’</td>
<td>8.83% 0 1 96,343</td>
<td>8.83%</td>
<td>8.98%</td>
<td>18.50%</td>
<td>7.87%</td>
<td>8.86%</td>
<td>20.30%</td>
<td>48.94%</td>
<td>21.88%</td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent stayed in hospital last year, other than childbirth, 0 otherwise. Taken from ‘hosp’ and ‘hospch’.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent had 6 of more visits to the GP in the last year, 0 otherwise. Taken from ‘hl2gp’.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent reports themselves as registered disabled (waves 6-8 and 10-15) or considers themselves disabled (wave 9). Taken from ‘hldsbl’ and ‘hldsbl1’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent reports health as poor or very poor, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent reports health as fair, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent reports health as good or very good, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummy 1 if respondent reports health as excellent, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marriage status</th>
<th>These dummy variables are created from the variable ‘mastat’.</th>
<th>Married</th>
<th>Divorced or separated</th>
<th>Co-habiting</th>
<th>Widowed</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>Dummy 1 if married, 0 otherwise.</td>
<td>53.31%</td>
<td>7.01%</td>
<td>12.07%</td>
<td>7.40%</td>
<td>20.22%</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>Dummy 1 if divorced or separated, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-habiting</td>
<td>Dummy 1 if co-habiting, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>Dummy 1 if widowed, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>Dummy 1 if never married, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>These dummy variables are created from the variable ‘age12’ which is the individuals age on 1st December of the year the fieldwork began.</th>
<th>Age 20 or less</th>
<th>Age 21 to 25</th>
<th>Age 26 to 30</th>
<th>Age 31 to 35</th>
<th>Age 36 to 40</th>
<th>Age 41 to 45</th>
<th>Age 46 to 50</th>
<th>Age 51 to 55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 20 or less</td>
<td>Dummy 1 if respondent’s age is 20 or less, 0 otherwise.</td>
<td>8.05%</td>
<td>8.38%</td>
<td>8.89%</td>
<td>9.58%</td>
<td>10.26%</td>
<td>8.94%</td>
<td>8.11%</td>
<td>8.11%</td>
</tr>
<tr>
<td>Age 21 to 25</td>
<td>Dummy 1 if respondent’s age is 21 to 25, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 26 to 30</td>
<td>Dummy 1 if respondent’s age is 26 to 30, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 31 to 35</td>
<td>Dummy 1 if respondent’s age is 31 to 35, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 36 to 40</td>
<td>Dummy 1 if respondent’s age is 36 to 40, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 41 to 45</td>
<td>Dummy 1 if respondent’s age is 41 to 45, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 46 to 50</td>
<td>Dummy 1 if respondent’s age is 46 to 50, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 51 to 55</td>
<td>Dummy 1 if respondent’s age is 51 to 55, 0 otherwise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Empirical differences between measures

| Age 56 to 60 | Dummy 1 if respondent’s age is 56 to 60, 0 otherwise. | 6.91% | 0 | 1 | 96,342 |
| Age 61 to 65 | Dummy 1 if respondent’s age is 61 to 65, 0 otherwise. | 5.58% | 0 | 1 | 96,342 |
| Age 66 to 70 | Dummy 1 if respondent’s age is 66 to 70, 0 otherwise. | 5.05% | 0 | 1 | 96,342 |
| Age 71 to 75 | Dummy 1 if respondent’s age is 71 to 75, 0 otherwise. | 4.87% | 0 | 1 | 96,342 |
| Age 76 to 80 | Dummy 1 if respondent’s age is 76 to 80, 0 otherwise. | 3.74% | 0 | 1 | 96,342 |
| Age 80 plus | Dummy 1 if respondent’s age is above 80, 0 otherwise. | 3.53% | 0 | 1 | 96,342 |
| **Gender** | Dummy 1 if male, 0 otherwise. From variable ‘sex’. | 46.00% | 0 | 1 | 96,343 |
| **Social relations** | Taken dummy variables are taken from the variables ‘frna’: “How often do you talk to any of your neighbours? Is it... on most days/once or twice a week/once or twice a month/less than once a month/never” and ‘frnb’: “How often do you meet friends or relatives who are not living with you? Is it... on most days/once or twice a week/once or twice a month/less than once a month/never” | |
| High contact friends and family | Dummy 1 if sees friends and family most days, 0 otherwise | 86.60% | 0 | 1 | 96,291 |
| High contact neighbours | Dummy 1 if sees neighbours most days, 0 otherwise | 76.98% | 0 | 1 | 96,292 |
| **Labour force status** | These dummy variables are created from the variable ‘jbstat’, which asks the individual to state the option which best describes their current activity. | |
| Employed | Dummy 1 if respondent employed, 0 otherwise | 51.53% | 0 | 1 | 96,321 |
| Long term sick | Dummy 1 if respondent on long term sick leave, 0 otherwise | 4.04% | 0 | 1 | 96,321 |
| Retired | Dummy 1 if respondent retired, 0 otherwise | 20.75% | 0 | 1 | 96,321 |
| Unemployed | Dummy 1 if respondent unemployed, 0 otherwise | 3.34% | 0 | 1 | 96,321 |
| Maternity leave | Dummy 1 if respondent on maternity leave, 0 otherwise | 0.41% | 0 | 1 | 96,321 |
| Self employed | Dummy 1 if respondent self-employed, 0 otherwise | 6.66% | 0 | 1 | 96,321 |
| Family carer | Dummy 1 if respondent employed, 0 otherwise | 6.82% | 0 | 1 | 96,321 |
| Student | Dummy 1 if respondent family carer, 0 otherwise | 5.83% | 0 | 1 | 96,321 |
| Other | Dummy 1 if respondent has other labour force status. Includes waiting to take up a job. This variable has also been recoded to include being on a Government training scheme. | 0.45% | 0 | 1 | 96,321 |
| **Time working** | Time spent on working type activity. Including paid hours (‘jbhrs’, ‘jbot’ and ‘jshrsl’), hours caring (‘aidhrs’), household work (‘howling’) and | 34.448 (21.178) | 0 | 171 | 95,229 |
commuting (‘jsttwt’ and ‘jbtwt’, adjusted to total hours).

### Housing

<table>
<thead>
<tr>
<th>Dummy</th>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Detached</strong></td>
<td>Dummy 1 if house type is detached, 0 otherwise. Taken from ‘hstype’</td>
<td>25.59%</td>
</tr>
<tr>
<td><strong>Semi</strong></td>
<td>Dummy 1 if house type is a semi detached, 0 otherwise.</td>
<td>32.05%</td>
</tr>
<tr>
<td><strong>Terrace</strong></td>
<td>Dummy 1 if house type is terraced, 0 otherwise.</td>
<td>26.72%</td>
</tr>
<tr>
<td><strong>Flat</strong></td>
<td>Dummy 1 if house type is a flat, 0 otherwise.</td>
<td>13.63%</td>
</tr>
<tr>
<td><strong>Other house</strong></td>
<td>Dummy 1 if house type classes as other, or shelter, 0 otherwise.</td>
<td>2.01%</td>
</tr>
<tr>
<td><strong>Shared garden</strong></td>
<td>Dummy 1 if respondent shares a garden, 0 otherwise. Taken from ‘hsgdns’.</td>
<td>6.61%</td>
</tr>
<tr>
<td><strong>Rented public</strong></td>
<td>Dummy 1 if house is rented from the council or housing association, 0 otherwise. Taken from the variable ‘tenure’.</td>
<td>8.39%</td>
</tr>
<tr>
<td><strong>Rented private</strong></td>
<td>Dummy 1 if house is rented from the private sector, 0 otherwise. Taken from the variable ‘tenure’.</td>
<td>8.39%</td>
</tr>
<tr>
<td><strong>Years at address</strong></td>
<td>Length of time lived in current address. Calculated from the date of interview ‘doiy4’ and year moved to present address ‘plnowy4’, allowing for non report of those who gave the response in previous waves.</td>
<td>11.490 (12.151)</td>
</tr>
<tr>
<td><strong>House value</strong></td>
<td>Taken from ‘hsval’, “About how much would you expect to get for your home if you sold it today?” Deflated using the RPI.</td>
<td>£53215.17</td>
</tr>
<tr>
<td><strong>Housing costs</strong></td>
<td>Net monthly housing costs. Taken from ‘xphsn’. Deflated using RPI.</td>
<td>£122.923</td>
</tr>
</tbody>
</table>

### Location

<table>
<thead>
<tr>
<th>Dummy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urban</strong></td>
<td>Dummy 1 if region classes as metropolitan, 0 otherwise.</td>
</tr>
<tr>
<td><strong>Regional dummies 1 to 18</strong></td>
<td>Standard UK regions distinguishing former metropolitan counties and inner and outer London.</td>
</tr>
</tbody>
</table>

### Neighbourhood

<table>
<thead>
<tr>
<th>Dummy</th>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Noisy neighbours</strong></td>
<td>Dummy 1 if reports problems with noisy neighbours, 0 otherwise. Taken from ‘hspbrh’.</td>
<td>11.00%</td>
</tr>
<tr>
<td><strong>Noisy street</strong></td>
<td>Dummy 1 if reports problems with noisy street, 0 otherwise. Taken from ‘hspbri’.</td>
<td>16.67%</td>
</tr>
<tr>
<td><strong>Crime and vandalism</strong></td>
<td>Dummy 1 if reports problems with crime and/or vandalism, 0 otherwise. Taken from ‘hspbrq’.</td>
<td>17.76%</td>
</tr>
<tr>
<td><strong>Pollution</strong></td>
<td>Dummy 1 if reports problems with pollution, grime or other environmental problems, 0 otherwise. Taken from ‘hspbrp’.</td>
<td>7.72%</td>
</tr>
<tr>
<td><strong>House prices</strong></td>
<td>Regional house price data from the Nationwide. Uses Q4 data for the year.</td>
<td>£53211 £29792 £126550</td>
</tr>
</tbody>
</table>
Not disaggregated by metropolitan area. Deflated by the RPI. (£19425)

Note: Excludes Northern Ireland.

Table 11: Measures of well-being, waves 7-14, unless stated

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
<th>Mean (sd)</th>
<th>Min</th>
<th>Max</th>
<th>Observations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent income, net current</td>
<td>Equivalised weekly net income. Equivalised using the modified OECD variable ‘eq_moecd’, which is taken from the additional net income files. Adjusted to 1987 prices using CHAW RPI from ONS.</td>
<td>£142.89 (94.80)</td>
<td>£-16.45</td>
<td>£3166.66</td>
<td>85,467</td>
</tr>
<tr>
<td>Equivalent consumption</td>
<td>Equivalised annual using ‘eq_moecd’, which is taken from the additional net income files. Adjusted to 1987 prices using CHAW RPI from ONS. Includes: Annual fuel costs (‘xpgasy’, ‘xplecy’, ‘xpoily’, ‘xpsfly’), aggregated net monthly housing costs (‘xphsn’), aggregate weekly expenditure on food (‘xfood’, taking mid point of the categories), cost of 12 consumer durables purchases in the last year (‘cd1cst’ to ‘cd12cst’), aggregate monthly health insurance, meals out and leisure (‘hcvrl’, ‘xmeal’, ‘xpleis’) summed to the household level and aggregate weekly child care (‘xchca’).</td>
<td>£2820.40 (1492.19)</td>
<td>£13386.43</td>
<td>£68019</td>
<td>57,226</td>
</tr>
<tr>
<td>SWB past</td>
<td>Taken from ‘qlfq’ and ‘qlfn’, recoded on 0 to 3 scale such that 3 is the most positive outcome. Aggregated.</td>
<td>4.79 (1.20)</td>
<td>0</td>
<td>6</td>
<td>13,074 (wave 11 only)</td>
</tr>
<tr>
<td>SWB future</td>
<td>Taken from ‘qlfs’ and ‘qlfj’, recoded on 0 to 3 scale such that 3 is the most positive outcome. Aggregated.</td>
<td>4.67 (1.17)</td>
<td>0</td>
<td>6</td>
<td>13,074 (wave 11 only)</td>
</tr>
<tr>
<td>GHQ Likert</td>
<td>Aggregate of the GHQ12 scores, based on the 0 to 3 scoring, taken from ‘hlgq1’, and reversed coded such that a higher score is less depression and anxiety.</td>
<td>24.85 (5.45)</td>
<td>0</td>
<td>36</td>
<td>94,149</td>
</tr>
<tr>
<td>GHQ Caseness</td>
<td>Aggregate of GHQ12, based on dichotomous scoring for each questions, taken from ‘hlgq2’, and reversed coded such that a higher score is less depression and anxiety.</td>
<td>10.11 (2.97)</td>
<td>0</td>
<td>12</td>
<td>94,149</td>
</tr>
<tr>
<td>GHQ Positive</td>
<td>This aggregates only the positively worded questions within the GHQ12</td>
<td>11.68 (2.44)</td>
<td>0</td>
<td>18</td>
<td>94,356</td>
</tr>
<tr>
<td>Measure</td>
<td>Description</td>
<td>Mean (SD)</td>
<td>Minimum</td>
<td>Maximum</td>
<td>N</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>GHQ Negative</td>
<td>This aggregates only the negatively worded questions within the GHQ12</td>
<td>13.16 (3.55)</td>
<td>0</td>
<td>18</td>
<td>94,327</td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>Taken from 'lfsato'</td>
<td>5.24 (1.29)</td>
<td>1</td>
<td>7</td>
<td>81,067 (ex. wave 11)</td>
</tr>
<tr>
<td>Domain weighted</td>
<td>Taken from 'lfsat1', 'lfsat2', 'lfsat3', 'lfsat4', 'lfsat5', 'lfsat6', 'lfsat7', 'lfsat8'. Weighted according to the method discussed in the text.</td>
<td>5.27 (0.838)</td>
<td>1.848</td>
<td>6.826</td>
<td>78,709 (ex. wave 11)</td>
</tr>
<tr>
<td>Sf-6d</td>
<td>Coding from an SPSS syntax file from John Brazier (30th July 2001) to derive the Sf-6d from the Sf-36 data.</td>
<td>0.833 (0.147)</td>
<td>0.296</td>
<td>1</td>
<td>25,670 (waves 9 &amp; 14 only)</td>
</tr>
</tbody>
</table>
APPENDIX 3.E

Table 12: OLS coefficients on various well-being measures, wave 11

<table>
<thead>
<tr>
<th></th>
<th>Equivalent Income</th>
<th>Equivalent Consump.</th>
<th>CASP-19</th>
<th>SWB Past</th>
<th>SWB Future</th>
<th>GHQ Likert</th>
<th>GHQ Caseness</th>
<th>GHQ Positive</th>
<th>GHQ Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income</td>
<td></td>
<td></td>
<td>0.020</td>
<td>0.025</td>
<td>0.025</td>
<td>0.005</td>
<td>0.003</td>
<td>0.001</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)**</td>
<td>(0.004)**</td>
<td>(0.004)**</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>No. children 0-4</td>
<td>-26.09</td>
<td>-347.22</td>
<td>-0.021</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(2.25)**</td>
<td>(34.07)**</td>
<td>(0.003)**</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>No. children 5-11</td>
<td>-19.34</td>
<td>-302.03</td>
<td>-0.011</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.005</td>
<td>0.013</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(1.51)**</td>
<td>(22.88)**</td>
<td>(0.002)**</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)**</td>
<td>(0.002)**</td>
<td>(0.003)</td>
</tr>
<tr>
<td>No. children 12-18</td>
<td>-21.60</td>
<td>-346.55</td>
<td>-0.017</td>
<td>-0.009</td>
<td>-0.012</td>
<td>-0.010</td>
<td>-0.012</td>
<td>-0.008</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(1.69)**</td>
<td>(25.52)**</td>
<td>(0.002)**</td>
<td>(0.004)</td>
<td>(0.004)**</td>
<td>(0.003)**</td>
<td>(0.005)</td>
<td>(0.003)**</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>Adults</td>
<td>28.38</td>
<td>-179.26</td>
<td>-0.018</td>
<td>-0.001</td>
<td>-0.006</td>
<td>0.003</td>
<td>0.005</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(4.69)**</td>
<td>(70.91)*</td>
<td>(0.007)**</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Adults squared</td>
<td>-4.40</td>
<td>10.51</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.83)**</td>
<td>(12.50)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>O level, or similar</td>
<td>13.54</td>
<td>105.68</td>
<td>0.011</td>
<td>0.017</td>
<td>0.013</td>
<td>-0.003</td>
<td>-0.012</td>
<td>-0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(2.53)**</td>
<td>(38.20)**</td>
<td>(0.004)**</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>A level or equivalent</td>
<td>22.99</td>
<td>316.34</td>
<td>0.013</td>
<td>0.018</td>
<td>0.019</td>
<td>-0.007</td>
<td>-0.024</td>
<td>-0.005</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(3.22)**</td>
<td>(48.66)**</td>
<td>(0.004)**</td>
<td>(0.008)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Degree or equivalent</td>
<td>47.74</td>
<td>562.82</td>
<td>0.021</td>
<td>0.035</td>
<td>0.026</td>
<td>-0.008</td>
<td>-0.031</td>
<td>-0.005</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(2.46)**</td>
<td>(37.19)**</td>
<td>(0.003)**</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.006)*</td>
</tr>
<tr>
<td>Problems walking</td>
<td>7.31</td>
<td>-40.73</td>
<td>-0.048</td>
<td>-0.017</td>
<td>-0.056</td>
<td>-0.033</td>
<td>-0.052</td>
<td>-0.034</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>(3.83)**</td>
<td>(57.83)</td>
<td>(0.005)**</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.011)</td>
<td>(0.006)**</td>
<td>(0.009)**</td>
</tr>
<tr>
<td>High load carer</td>
<td>-12.10</td>
<td>-232.40</td>
<td>-0.057</td>
<td>-0.045</td>
<td>-0.072</td>
<td>-0.050</td>
<td>-0.087</td>
<td>-0.025</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>(6.33)**</td>
<td>(95.60)*</td>
<td>(0.009)**</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.011)</td>
<td>(0.018)</td>
<td>(0.010)*</td>
<td>(0.014)**</td>
</tr>
<tr>
<td>Hospital visit</td>
<td>2.98</td>
<td>1.86</td>
<td>-0.003</td>
<td>-0.003</td>
<td>0.001</td>
<td>-0.010</td>
<td>-0.021</td>
<td>-0.014</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(3.07)**</td>
<td>(46.33)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.005)**</td>
<td>(0.007)</td>
</tr>
<tr>
<td>High GP visits</td>
<td>-1.72</td>
<td>-17.68</td>
<td>-0.002</td>
<td>-0.003</td>
<td>0.005</td>
<td>-0.019</td>
<td>-0.046</td>
<td>-0.013</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>(2.42)**</td>
<td>(36.50)</td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)**</td>
<td>(0.005)**</td>
</tr>
<tr>
<td>SRH poor</td>
<td>-5.79</td>
<td>-2.79</td>
<td>-0.101</td>
<td>-0.101</td>
<td>-0.134</td>
<td>-0.128</td>
<td>-0.226</td>
<td>-0.109</td>
<td>-0.147</td>
</tr>
<tr>
<td></td>
<td>(3.69)</td>
<td>(55.84)</td>
<td>(0.005)**</td>
<td>(0.009)</td>
<td>(0.008)**</td>
<td>(0.006)**</td>
<td>(0.011)**</td>
<td>(0.006)**</td>
<td>(0.008)**</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>t-Statistic</td>
<td>p-Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRH fair</td>
<td>13.59</td>
<td>1.03</td>
<td>12.94</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRH excellent</td>
<td>10.13</td>
<td>0.47</td>
<td>20.47</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>-17.51</td>
<td>0.20</td>
<td>-20.13</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohabiting</td>
<td>7.28</td>
<td>0.00</td>
<td>72.87</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>7.59</td>
<td>0.01</td>
<td>75.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>1.53</td>
<td>0.01</td>
<td>15.29</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 21 to 25</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 26 to 30</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 31 to 35</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 36 to 40</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 41 to 45</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 46 to 50</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 51 to 55</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 56 to 60</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 61 to 65</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 66 to 70</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 71 to 75</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 76 to 80</td>
<td>11.90</td>
<td>0.02</td>
<td>118.90</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 80 plus</td>
<td>Male</td>
<td>(8.43)*</td>
<td>(127.36)**</td>
<td>(0.012)*</td>
<td>(0.020)**</td>
<td>(0.019)</td>
<td>(0.014)*</td>
<td>(0.024)</td>
<td>(0.013)**</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>---------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>-17.63</td>
<td>-0.12</td>
<td>-39.89</td>
<td>-0.010</td>
<td>-0.026</td>
<td>-0.024</td>
<td>0.028</td>
<td>0.035</td>
<td>0.012</td>
<td>0.043</td>
</tr>
<tr>
<td>(1.73)</td>
<td></td>
<td>(26.14)</td>
<td>(0.002)**</td>
<td>(0.004)**</td>
<td>(0.004)**</td>
<td>(0.003)**</td>
<td>(0.005)**</td>
<td>(0.003)**</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>High contact friends/fam</td>
<td>-3.83</td>
<td>(2.49)</td>
<td>(37.68)</td>
<td>(0.003)**</td>
<td>(0.006)**</td>
<td>(0.006)**</td>
<td>(0.004)**</td>
<td>(0.007)**</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>(2.03)**</td>
<td></td>
<td>(30.66)**</td>
<td>(0.003)**</td>
<td>(0.005)**</td>
<td>(0.005)**</td>
<td>(0.003)**</td>
<td>(0.006)**</td>
<td>(0.003)**</td>
<td>(0.005)**</td>
</tr>
<tr>
<td>High contact neighbours</td>
<td>-9.83</td>
<td>(4.13)**</td>
<td>(62.43)**</td>
<td>(0.006)*</td>
<td>(0.010)</td>
<td>(0.009)**</td>
<td>(0.007)</td>
<td>(0.012)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>(9.08)**</td>
<td></td>
<td>(76.78)**</td>
<td>(0.007)**</td>
<td>(0.012)**</td>
<td>(0.012)**</td>
<td>(0.009)**</td>
<td>(0.015)**</td>
<td>(0.008)**</td>
<td>(0.012)**</td>
</tr>
<tr>
<td>Maternity leave</td>
<td>13.42</td>
<td>(13.98)</td>
<td>(211.30)</td>
<td>(0.019)</td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.024)</td>
<td>(0.040)</td>
<td>(0.022)*</td>
</tr>
<tr>
<td>(3.77)</td>
<td></td>
<td>(50.86)**</td>
<td>(0.005)**</td>
<td>(0.008)**</td>
<td>(0.008)**</td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.005)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Retired</td>
<td>-43.41</td>
<td>-47.50</td>
<td>0.012</td>
<td>0.010</td>
<td>0.026</td>
<td>0.013</td>
<td>0.014</td>
<td>0.005</td>
<td>0.022</td>
</tr>
<tr>
<td>(4.13)**</td>
<td></td>
<td>(62.43)**</td>
<td>(0.006)*</td>
<td>(0.010)</td>
<td>(0.009)**</td>
<td>(0.007)</td>
<td>(0.012)</td>
<td>(0.006)</td>
<td>(0.009)*</td>
</tr>
<tr>
<td>Unemployed</td>
<td>-71.80</td>
<td>-880.61</td>
<td>-0.035</td>
<td>-0.057</td>
<td>-0.045</td>
<td>-0.060</td>
<td>-0.098</td>
<td>-0.052</td>
<td>-0.068</td>
</tr>
<tr>
<td>(5.08)**</td>
<td></td>
<td>(76.78)**</td>
<td>(0.007)**</td>
<td>(0.012)**</td>
<td>(0.012)**</td>
<td>(0.009)**</td>
<td>(0.015)**</td>
<td>(0.008)**</td>
<td>(0.012)**</td>
</tr>
<tr>
<td>Student</td>
<td>-26.12</td>
<td>-45.80</td>
<td>0.021</td>
<td>0.055</td>
<td>-0.002</td>
<td>-0.015</td>
<td>0.007</td>
<td>-0.013</td>
<td></td>
</tr>
<tr>
<td>(4.93)**</td>
<td></td>
<td>(74.45)**</td>
<td>(0.007)**</td>
<td>(0.012)**</td>
<td>(0.011)**</td>
<td>(0.009)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Other activity</td>
<td>-40.23</td>
<td>-352.32</td>
<td>0.036</td>
<td>0.017</td>
<td>0.050</td>
<td>0.004</td>
<td>-0.025</td>
<td>0.013</td>
<td>-0.004</td>
</tr>
<tr>
<td>(10.87)**</td>
<td></td>
<td>(164.21)*</td>
<td>(0.015)*</td>
<td>(0.025)</td>
<td>(0.025)*</td>
<td>(0.019)</td>
<td>(0.031)</td>
<td>(0.017)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Constant</td>
<td>129.59</td>
<td>3950.40</td>
<td>0.622</td>
<td>0.624</td>
<td>0.616</td>
<td>0.679</td>
<td>0.830</td>
<td>0.666</td>
<td>0.690</td>
</tr>
<tr>
<td>(9.89)**</td>
<td></td>
<td>(149.49)**</td>
<td>(0.017)**</td>
<td>(0.028)**</td>
<td>(0.027)**</td>
<td>(0.021)**</td>
<td>(0.034)**</td>
<td>(0.019)**</td>
<td>(0.027)**</td>
</tr>
<tr>
<td>Adj R-squared</td>
<td>0.2217</td>
<td>0.2783</td>
<td>0.2573</td>
<td>0.1456</td>
<td>0.1843</td>
<td>0.1811</td>
<td>0.1699</td>
<td>0.1557</td>
<td>0.1692</td>
</tr>
<tr>
<td>Observations</td>
<td>9531</td>
<td>9531</td>
<td>9479</td>
<td>9479</td>
<td>9479</td>
<td>9479</td>
<td>9478</td>
<td>9478</td>
<td></td>
</tr>
</tbody>
</table>

F(43,9487) = 64.14  F(43,9487) = 86.46  F(44,9434) = 74.29  F(44,9434) = 37.70  F(44,9434) = 38.53  F(44,9434) = 48.63  F(44,9434) = 45.10  F(44,9433) = 40.71  F(44,9433) = 44.87
Table 13: OLS coefficients on various well-being measures, waves 9 & 14

<table>
<thead>
<tr>
<th></th>
<th>Equivalent Income</th>
<th>Equivalent Consump.</th>
<th>Life satisfaction</th>
<th>Domain weighted</th>
<th>GHQ Likert</th>
<th>GHQ Caseness</th>
<th>GHQ Positive</th>
<th>GHQ Negative</th>
<th>SF-6d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income</td>
<td>0.018</td>
<td>0.016</td>
<td>0.007</td>
<td>0.006</td>
<td>0.003</td>
<td>0.010</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)**</td>
<td>(0.002)**</td>
<td>(0.002)**</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)**</td>
<td>(0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. children 0-4</td>
<td>-25.75</td>
<td>-281.87</td>
<td>0.004</td>
<td>-0.023</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(1.93)**</td>
<td>(37.90)**</td>
<td>(0.005)</td>
<td>(0.003)**</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>No. children 5-11</td>
<td>-20.59</td>
<td>-292.91</td>
<td>-0.009</td>
<td>-0.016</td>
<td>0.000</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(1.56)**</td>
<td>(25.23)**</td>
<td>(0.002)**</td>
<td>(0.003)**</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>No. children 12-18</td>
<td>-20.66</td>
<td>-388.04</td>
<td>-0.015</td>
<td>-0.010</td>
<td>-0.005</td>
<td>-0.007</td>
<td>-0.002</td>
<td>-0.008</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(1.60)**</td>
<td>(23.71)**</td>
<td>(0.004)**</td>
<td>(0.002)**</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.004)**</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>28.78</td>
<td>-116.26</td>
<td>-0.021</td>
<td>0.000</td>
<td>-0.003</td>
<td>-0.009</td>
<td>-0.009</td>
<td>0.003</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(4.02)**</td>
<td>(69.67)**</td>
<td>(0.010)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.012)</td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Adults squared</td>
<td>-4.39</td>
<td>-4.88</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.66)**</td>
<td>(11.44)**</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>O level, or similar</td>
<td>13.39</td>
<td>85.34</td>
<td>-0.015</td>
<td>-0.004</td>
<td>0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.006</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(1.93)**</td>
<td>(30.60)**</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>A level or equivalent</td>
<td>25.11</td>
<td>343.38</td>
<td>-0.025</td>
<td>-0.011</td>
<td>-0.001</td>
<td>-0.011</td>
<td>-0.002</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(2.73)**</td>
<td>(44.69)**</td>
<td>(0.007)**</td>
<td>(0.005)**</td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Degree or equivalent</td>
<td>46.58</td>
<td>529.75</td>
<td>-0.026</td>
<td>-0.014</td>
<td>-0.001</td>
<td>-0.017</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(2.13)**</td>
<td>(35.75)**</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.006)**</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Problems walking</td>
<td>-1.97</td>
<td>-14.61</td>
<td>-0.045</td>
<td>-0.034</td>
<td>-0.025</td>
<td>-0.044</td>
<td>-0.029</td>
<td>-0.021</td>
<td>-0.116</td>
</tr>
<tr>
<td></td>
<td>(2.04)**</td>
<td>(31.79)**</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.008)**</td>
<td>(0.004)**</td>
<td>(0.006)**</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>High load carer</td>
<td>-7.81</td>
<td>-138.47</td>
<td>-0.045</td>
<td>-0.017</td>
<td>-0.031</td>
<td>-0.061</td>
<td>-0.011</td>
<td>-0.051</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(3.86)**</td>
<td>(79.57)**</td>
<td>(0.019)</td>
<td>(0.012)</td>
<td>(0.011)**</td>
<td>(0.020)**</td>
<td>(0.009)</td>
<td>(0.015)**</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Hospital visit</td>
<td>8.49</td>
<td>115.96</td>
<td>-0.013</td>
<td>-0.009</td>
<td>-0.015</td>
<td>-0.033</td>
<td>-0.017</td>
<td>-0.012</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>(2.99)**</td>
<td>(39.28)**</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.009)**</td>
<td>(0.005)**</td>
<td>(0.007)</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>High GP visits</td>
<td>-2.17</td>
<td>-35.33</td>
<td>-0.017</td>
<td>-0.016</td>
<td>-0.029</td>
<td>-0.054</td>
<td>-0.020</td>
<td>-0.038</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>(1.95)**</td>
<td>(32.59)**</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.007)**</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.003)**</td>
</tr>
<tr>
<td>SRH poor</td>
<td>-13.05</td>
<td>-170.73</td>
<td>-0.174</td>
<td>-0.124</td>
<td>-0.130</td>
<td>-0.218</td>
<td>-0.111</td>
<td>-0.149</td>
<td>-0.156</td>
</tr>
<tr>
<td></td>
<td>(2.58)**</td>
<td>(49.17)**</td>
<td>(0.010)</td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.013)**</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.005)**</td>
</tr>
<tr>
<td>SRH fair</td>
<td>-11.08</td>
<td>-152.26</td>
<td>-0.081</td>
<td>-0.067</td>
<td>-0.046</td>
<td>-0.071</td>
<td>-0.033</td>
<td>-0.059</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>(1.70)**</td>
<td>(30.46)**</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.003)**</td>
</tr>
<tr>
<td></td>
<td>SRH excellent</td>
<td>Divorced or separated</td>
<td>Cohabiting</td>
<td>Widowed</td>
<td>Never married</td>
<td>Age 21 to 25</td>
<td>Age 26 to 30</td>
<td>Age 31 to 35</td>
<td>Age 36 to 40</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>---------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>73.25</td>
<td>(34.73)*</td>
<td></td>
<td>-293.08</td>
<td>-87.47</td>
<td>-36.45</td>
<td>-403.85</td>
<td>104.32</td>
<td>213.19</td>
</tr>
<tr>
<td></td>
<td>0.062</td>
<td>(0.004)**</td>
<td></td>
<td>-0.097</td>
<td>-0.044</td>
<td>-0.064</td>
<td>-0.062</td>
<td>-0.071</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>0.053</td>
<td>(0.003)**</td>
<td></td>
<td>-0.080</td>
<td>-0.036</td>
<td>-0.050</td>
<td>-0.048</td>
<td>-0.055</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td>0.039</td>
<td>(0.003)**</td>
<td></td>
<td>-0.024</td>
<td>-0.020</td>
<td>-0.028</td>
<td>-0.001</td>
<td>-0.027</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>0.019</td>
<td>(0.004)**</td>
<td></td>
<td>-0.052</td>
<td>-0.014</td>
<td>-0.011</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>0.060</td>
<td>(0.004)**</td>
<td></td>
<td>-0.015</td>
<td>-0.015</td>
<td>-0.032</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>(0.002)**</td>
<td></td>
<td>-0.033</td>
<td>-0.024</td>
<td>-0.016</td>
<td>0.002</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

**Empirical differences between measures**
### Empirical differences between measures

<table>
<thead>
<tr>
<th>Group</th>
<th>Diff</th>
<th>Standard Error</th>
<th>t-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.73</td>
<td>-32.85</td>
<td>0.007</td>
<td>0.028</td>
</tr>
<tr>
<td>High contact friends/fam</td>
<td>-5.41</td>
<td>16.35</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>High contact neighbours</td>
<td>-8.70</td>
<td>-186.89</td>
<td>0.034</td>
<td>0.025</td>
</tr>
<tr>
<td>Long term sick</td>
<td>-45.54</td>
<td>-786.53</td>
<td>-0.47</td>
<td>-0.31</td>
</tr>
<tr>
<td>Retired</td>
<td>-39.70</td>
<td>87.55</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>Unemployed</td>
<td>-66.98</td>
<td>-803.98</td>
<td>-0.76</td>
<td>-0.39</td>
</tr>
<tr>
<td>Maternity leave</td>
<td>17.30</td>
<td>218.28</td>
<td>0.023</td>
<td>-0.03</td>
</tr>
<tr>
<td>Self employed</td>
<td>0.26</td>
<td>324.48</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>Family carer</td>
<td>-26.50</td>
<td>-463.97</td>
<td>0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td>Student</td>
<td>-31.99</td>
<td>59.92</td>
<td>0.004</td>
<td>-0.016</td>
</tr>
<tr>
<td>Other activity</td>
<td>-45.22</td>
<td>-652.01</td>
<td>-0.35</td>
<td>-0.03</td>
</tr>
<tr>
<td>Wave 9</td>
<td>-10.12</td>
<td>-253.54</td>
<td>-0.006</td>
<td>-0.005</td>
</tr>
<tr>
<td>Constant</td>
<td>129.29</td>
<td>3991.01</td>
<td>0.710</td>
<td>0.677</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.2064</td>
<td>0.2587</td>
<td>0.2017</td>
<td>0.2934</td>
</tr>
<tr>
<td>Observations</td>
<td>13108</td>
<td>13108</td>
<td>13034</td>
<td>13034</td>
</tr>
</tbody>
</table>

**Notes:**
- **F**: F(44,10560) = 88.99
- **Constant**: F(8,1308) = 129.29
- **R-squared**: F(45, 10513) = 309.13
### Table 14: OLS-FE coefficients on various well-being measures, waves 9 & 14

<table>
<thead>
<tr>
<th>Measure</th>
<th>Equivalent Income</th>
<th>Equivalent Consump.</th>
<th>Life satisfaction</th>
<th>Domain</th>
<th>GHQ Likert</th>
<th>GHQ Caseness</th>
<th>GHQ Positive</th>
<th>GHQ Negative</th>
<th>SF-6d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income</td>
<td>0.017</td>
<td>0.010</td>
<td>0.003</td>
<td>0.005</td>
<td>-0.001</td>
<td>0.007</td>
<td>-0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)**</td>
<td>(0.002)**</td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. children 0-4</td>
<td>-21.3</td>
<td>-217.9</td>
<td>-0.003</td>
<td>-0.019</td>
<td>-0.006</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.009</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(2.96)**</td>
<td>(87.81)**</td>
<td>(0.006)</td>
<td>(0.003)**</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>No. children 5-11</td>
<td>-9.8</td>
<td>-288.1</td>
<td>-0.002</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.009</td>
<td>0.002</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(2.38)**</td>
<td>(70.90)**</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>No. children 12-18</td>
<td>-12.1</td>
<td>-327.9</td>
<td>-0.005</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.006</td>
<td>-0.001</td>
<td>-0.007</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(2.54)**</td>
<td>(78.79)**</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Adults</td>
<td>13.1</td>
<td>-32.5</td>
<td>-0.027</td>
<td>-0.005</td>
<td>0.004</td>
<td>0.010</td>
<td>0.000</td>
<td>0.009</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(7.04)</td>
<td>(169.07)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.018)</td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Adults squared</td>
<td>-1.4</td>
<td>-17.5</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(27.51)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>O level, or similar</td>
<td>6.6</td>
<td>17.9</td>
<td>0.002</td>
<td>0.006</td>
<td>0.006</td>
<td>0.008</td>
<td>0.007</td>
<td>0.006</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(11.46)</td>
<td>(311.57)</td>
<td>(0.023)</td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.029)</td>
<td>(0.016)</td>
<td>(0.021)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>A level or equivalent</td>
<td>3.6</td>
<td>118.6</td>
<td>0.011</td>
<td>0.011</td>
<td>0.001</td>
<td>-0.006</td>
<td>0.007</td>
<td>-0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(11.94)</td>
<td>(322.96)</td>
<td>(0.024)</td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.031)</td>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Degree or equivalent</td>
<td>1.9</td>
<td>-7.2</td>
<td>0.008</td>
<td>0.010</td>
<td>0.010</td>
<td>0.016</td>
<td>0.012</td>
<td>0.008</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(10.85)</td>
<td>(295.77)</td>
<td>(0.021)</td>
<td>(0.012)</td>
<td>(0.016)</td>
<td>(0.028)</td>
<td>(0.016)</td>
<td>(0.019)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Problems walking</td>
<td>-5.5</td>
<td>-3.2</td>
<td>-0.040</td>
<td>-0.028</td>
<td>-0.026</td>
<td>-0.045</td>
<td>-0.025</td>
<td>-0.027</td>
<td>-0.084</td>
</tr>
<tr>
<td></td>
<td>(4.29)</td>
<td>(102.64)</td>
<td>(0.009)</td>
<td>(0.005)**</td>
<td>(0.006)**</td>
<td>(0.011)**</td>
<td>(0.006)**</td>
<td>(0.008)**</td>
<td>(0.005)**</td>
</tr>
<tr>
<td>High load carer</td>
<td>0.4</td>
<td>-76.3</td>
<td>-0.035</td>
<td>-0.022</td>
<td>-0.024</td>
<td>-0.055</td>
<td>-0.019</td>
<td>-0.028</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(9.57)</td>
<td>(256.14)</td>
<td>(0.019)</td>
<td>(0.012)*</td>
<td>(0.014)</td>
<td>(0.025)*</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Hospital visit</td>
<td>3.0</td>
<td>46.7</td>
<td>0.000</td>
<td>0.003</td>
<td>-0.002</td>
<td>-0.006</td>
<td>-0.004</td>
<td>-0.001</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>(3.51)</td>
<td>(91.16)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.004)**</td>
</tr>
<tr>
<td>High GP visits</td>
<td>5.4</td>
<td>-103.6</td>
<td>-0.007</td>
<td>-0.011</td>
<td>-0.024</td>
<td>-0.042</td>
<td>-0.020</td>
<td>-0.028</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(3.03)</td>
<td>(80.60)</td>
<td>(0.006)</td>
<td>(0.003)**</td>
<td>(0.004)**</td>
<td>(0.008)**</td>
<td>(0.004)**</td>
<td>(0.005)**</td>
<td>(0.003)**</td>
</tr>
<tr>
<td>SRH poor</td>
<td>-4.4</td>
<td>55.8</td>
<td>-0.103</td>
<td>-0.077</td>
<td>-0.106</td>
<td>-0.176</td>
<td>-0.101</td>
<td>-0.111</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>(5.10)</td>
<td>(128.90)</td>
<td>(0.010)**</td>
<td>(0.006)**</td>
<td>(0.008)**</td>
<td>(0.013)**</td>
<td>(0.007)**</td>
<td>(0.009)**</td>
<td>(0.006)**</td>
</tr>
<tr>
<td>SRH fair</td>
<td>-1.1</td>
<td>42.0</td>
<td>-0.046</td>
<td>-0.035</td>
<td>-0.033</td>
<td>-0.058</td>
<td>-0.028</td>
<td>-0.038</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>(2.96)</td>
<td>(77.68)</td>
<td>(0.006)**</td>
<td>(0.003)**</td>
<td>(0.004)**</td>
<td>(0.008)</td>
<td>(0.004)**</td>
<td>(0.005)**</td>
<td>(0.003)**</td>
</tr>
</tbody>
</table>
### Chapter 3

#### Empirical differences between measures

<table>
<thead>
<tr>
<th>SRH excellent</th>
<th>Divorced or separated</th>
<th>Cohabiting</th>
<th>Widowed</th>
<th>Never married</th>
<th>Age 21 to 25</th>
<th>Age 26 to 30</th>
<th>Age 31 to 35</th>
<th>Age 36 to 40</th>
<th>Age 41 to 45</th>
<th>Age 46 to 50</th>
<th>Age 51 to 55</th>
<th>Age 56 to 60</th>
<th>Age 61 to 65</th>
<th>Age 66 to 70</th>
<th>Age 71 to 75</th>
<th>Age 76 to 80</th>
<th>Age 80 plus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>-13.9</td>
<td>1.8</td>
<td>6.2</td>
<td>-14.6</td>
<td>-22.2</td>
<td>-13.4</td>
<td>-11.3</td>
<td>-9.0</td>
<td>-8.5</td>
<td>-1.8</td>
<td>-9.6</td>
<td>-14.6</td>
<td>-12.2</td>
<td>-13.5</td>
<td>-30.8</td>
<td>-32.3</td>
<td>-32.9</td>
</tr>
<tr>
<td>(2.90)</td>
<td>(78.12)</td>
<td>(7.14)</td>
<td>(4.96)</td>
<td>(7.23)*</td>
<td>(32.85)</td>
<td>(23.88)</td>
<td>(34.68)</td>
<td>(45.52)*</td>
<td>(56.41)*</td>
<td>(67.36)</td>
<td>(25.52)</td>
<td>(89.31)</td>
<td>(100.3)</td>
<td>(111.3)</td>
<td>(122.3)</td>
<td>(133.3)</td>
<td>(144.3)</td>
</tr>
<tr>
<td>(0.006)**</td>
<td>(0.014)**</td>
<td>(0.026)</td>
<td>(0.069)</td>
<td>(0.014)*</td>
<td>(0.112)</td>
<td>(0.134)</td>
<td>(0.156)</td>
<td>(0.178)</td>
<td>(0.189)</td>
<td>(0.237)</td>
<td>(0.150)</td>
<td>(0.178)</td>
<td>(0.200)*</td>
<td>(0.223)*</td>
<td>(0.244)*</td>
<td>(0.266)*</td>
<td>(0.287)*</td>
</tr>
<tr>
<td>(0.003)**</td>
<td>(0.008)**</td>
<td>(0.015)</td>
<td>(0.039)</td>
<td>(0.001)*</td>
<td>(0.064)</td>
<td>(0.076)</td>
<td>(0.089)*</td>
<td>(0.101)*</td>
<td>(0.112)*</td>
<td>(0.142)</td>
<td>(0.185)</td>
<td>(0.230)</td>
<td>(0.273)</td>
<td>(0.314)</td>
<td>(0.350)</td>
<td>(0.374)</td>
<td>(0.399)</td>
</tr>
<tr>
<td>(0.004)**</td>
<td>(0.012)</td>
<td>(0.126)</td>
<td>(0.150)</td>
<td>(0.131)</td>
<td>(0.164)</td>
<td>(0.181)</td>
<td>(0.195)</td>
<td>(0.215)</td>
<td>(0.223)</td>
<td>(0.257)</td>
<td>(0.273)</td>
<td>(0.299)</td>
<td>(0.343)</td>
<td>(0.361)</td>
<td>(0.397)</td>
<td>(0.425)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>(0.007)</td>
<td>(0.019)</td>
<td>(0.035)</td>
<td>(0.100)</td>
<td>(0.129)</td>
<td>(0.161)</td>
<td>(0.181)</td>
<td>(0.195)</td>
<td>(0.215)</td>
<td>(0.223)</td>
<td>(0.259)</td>
<td>(0.273)</td>
<td>(0.309)</td>
<td>(0.345)</td>
<td>(0.361)</td>
<td>(0.397)</td>
<td>(0.425)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>(0.004)**</td>
<td>(0.019)</td>
<td>(0.035)</td>
<td>(0.100)</td>
<td>(0.129)</td>
<td>(0.161)</td>
<td>(0.181)</td>
<td>(0.195)</td>
<td>(0.215)</td>
<td>(0.223)</td>
<td>(0.259)</td>
<td>(0.273)</td>
<td>(0.309)</td>
<td>(0.345)</td>
<td>(0.361)</td>
<td>(0.397)</td>
<td>(0.425)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>(0.005)**</td>
<td>(0.019)</td>
<td>(0.035)</td>
<td>(0.100)</td>
<td>(0.129)</td>
<td>(0.161)</td>
<td>(0.181)</td>
<td>(0.195)</td>
<td>(0.215)</td>
<td>(0.223)</td>
<td>(0.259)</td>
<td>(0.273)</td>
<td>(0.309)</td>
<td>(0.345)</td>
<td>(0.361)</td>
<td>(0.397)</td>
<td>(0.425)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>(0.003)**</td>
<td>(0.019)</td>
<td>(0.035)</td>
<td>(0.100)</td>
<td>(0.129)</td>
<td>(0.161)</td>
<td>(0.181)</td>
<td>(0.195)</td>
<td>(0.215)</td>
<td>(0.223)</td>
<td>(0.259)</td>
<td>(0.273)</td>
<td>(0.309)</td>
<td>(0.345)</td>
<td>(0.361)</td>
<td>(0.397)</td>
<td>(0.425)</td>
<td>(0.462)</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>Wave 9</td>
<td>Long term sick</td>
<td>Retired</td>
<td>Unemployed</td>
<td>Maternity leave</td>
<td>Self employed</td>
<td>Family carer</td>
<td>Student</td>
<td>Other activity</td>
<td>Wave 9</td>
<td>Constant</td>
<td>R squared within</td>
<td>Observations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------</td>
<td>--------</td>
<td>----------------</td>
<td>---------</td>
<td>------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------</td>
<td>---------------</td>
<td>--------</td>
<td>----------</td>
<td>----------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High contact friends/fam</td>
<td>1.0</td>
<td>-9.9</td>
<td>0.015</td>
<td>0.015</td>
<td>0.008</td>
<td>0.012</td>
<td>0.005</td>
<td>0.011</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High contact neighbours</td>
<td>-3.9</td>
<td>-55.6</td>
<td>0.020</td>
<td>0.011</td>
<td>0.010</td>
<td>0.018</td>
<td>0.008</td>
<td>0.012</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long term sick</td>
<td>-42.0</td>
<td>-203.8</td>
<td>-0.069</td>
<td>-0.032</td>
<td>-0.017</td>
<td>-0.032</td>
<td>-0.025</td>
<td>-0.010</td>
<td>-0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>-50.0</td>
<td>-287.6</td>
<td>0.009</td>
<td>0.014</td>
<td>0.006</td>
<td>0.002</td>
<td>0.002</td>
<td>0.010</td>
<td>-0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>-44.9</td>
<td>-121.7</td>
<td>-0.046</td>
<td>-0.031</td>
<td>-0.037</td>
<td>-0.066</td>
<td>-0.032</td>
<td>-0.041</td>
<td>-0.025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternity leave</td>
<td>-1.9</td>
<td>-41.4</td>
<td>0.017</td>
<td>0.009</td>
<td>0.024</td>
<td>0.014</td>
<td>0.016</td>
<td>0.032</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self employed</td>
<td>-6.9</td>
<td>135.5</td>
<td>-0.010</td>
<td>0.001</td>
<td>-0.010</td>
<td>-0.016</td>
<td>-0.012</td>
<td>-0.008</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family carer</td>
<td>-37.0</td>
<td>-189.9</td>
<td>-0.009</td>
<td>-0.012</td>
<td>0.005</td>
<td>0.019</td>
<td>0.005</td>
<td>0.005</td>
<td>-0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>-50.0</td>
<td>-167.9</td>
<td>0.020</td>
<td>-0.004</td>
<td>0.023</td>
<td>0.035</td>
<td>0.022</td>
<td>0.023</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other activity</td>
<td>-44.6</td>
<td>-275.1</td>
<td>0.035</td>
<td>-0.002</td>
<td>-0.008</td>
<td>-0.042</td>
<td>0.000</td>
<td>-0.017</td>
<td>-0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 9</td>
<td>-10.4</td>
<td>45.0</td>
<td>0.046</td>
<td>0.033</td>
<td>0.015</td>
<td>0.035</td>
<td>0.015</td>
<td>0.014</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>172.7</td>
<td>1325.6</td>
<td>0.404</td>
<td>0.502</td>
<td>0.581</td>
<td>0.540</td>
<td>0.554</td>
<td>0.608</td>
<td>0.737</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R squared within</td>
<td>0.0541</td>
<td>0.0493</td>
<td>0.0608</td>
<td>0.0993</td>
<td>0.0720</td>
<td>0.0671</td>
<td>0.0635</td>
<td>0.0585</td>
<td>0.2150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>20487</td>
<td>13199</td>
<td>20358</td>
<td>20358</td>
<td>20358</td>
<td>20358</td>
<td>20358</td>
<td>20358</td>
<td>20207</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>13579</td>
<td>10625</td>
<td>13518</td>
<td>13518</td>
<td>13518</td>
<td>13518</td>
<td>13518</td>
<td>13518</td>
<td>13457</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: standard errors in brackets; ** is p < 0.01, * is p < 0.05; household income is net, current income.
### APPENDIX 3.F

Table 15: Studies finding a positive impact of education on SWB

<table>
<thead>
<tr>
<th>Study</th>
<th>Dataset</th>
<th>Well-being question</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alesina et al. (2004); Blanchflower &amp; Oswald (1997 &amp; 2004); Di Tella et al. (2003); Easterlin (2001); Graham &amp; Pettinato (2001)</td>
<td>US, GSS</td>
<td>General happiness (1-3)</td>
<td>Life satisfaction (1-3)</td>
</tr>
<tr>
<td>Alesina et al. (2004); Blanchflower &amp; Oswald (1997 &amp; 2004); Di Tella et al. (2003); Easterlin (2001); Graham &amp; Pettinato (2001)</td>
<td>US, GSS</td>
<td>General happiness (1-3)</td>
<td>Life satisfaction (1-3)</td>
</tr>
<tr>
<td>Bukenya et al. (2003)</td>
<td>US</td>
<td>Personal satisfaction (1-3)</td>
<td>Life satisfaction (1-3)</td>
</tr>
<tr>
<td>Smith (2003)</td>
<td>Russia, WVS</td>
<td>General happiness (1-4)</td>
<td>Life satisfaction (1-4)</td>
</tr>
<tr>
<td>Blanchflower &amp; Oswald (2004)</td>
<td>ISSP</td>
<td>General happiness (1-4)</td>
<td>Life satisfaction (1-4)</td>
</tr>
<tr>
<td>Dorn et al. (2007)</td>
<td>Eurobarometer</td>
<td>Life satisfaction (1-4)</td>
<td>Life satisfaction (1-4)</td>
</tr>
<tr>
<td>Di Tella et al. (2001, 2003); Hudson (2006); Alesina et al. (2004); Blanchflower &amp; Oswald (1997 &amp; 2004)</td>
<td>Eurobarometer</td>
<td>Life satisfaction (1-4)</td>
<td>Life satisfaction (1-4)</td>
</tr>
<tr>
<td>Subramanian et al. (2005)</td>
<td>US, SCB</td>
<td>Life satisfaction (1-4)</td>
<td>Life satisfaction (1-4)</td>
</tr>
<tr>
<td>Fahey &amp; Smyth (2005)</td>
<td>European, EVS</td>
<td>Life satisfaction (1-10)</td>
<td>Life satisfaction (1-10)</td>
</tr>
<tr>
<td>Frey &amp; Stutzer (2000); Stutzer (2004)</td>
<td>Switzerland</td>
<td>Life satisfaction (1-10)</td>
<td>Life satisfaction (1-10)</td>
</tr>
<tr>
<td>Heady &amp; Wooden (2004)</td>
<td>Germany, GSOEP</td>
<td>Life satisfaction (0-10)</td>
<td>Life satisfaction (0-10)</td>
</tr>
<tr>
<td>Hurtog &amp; Oosterbeek (1998)</td>
<td>Netherlands</td>
<td>Best possible life (1-10)</td>
<td>Best possible life (1-10)</td>
</tr>
<tr>
<td>Schwarze &amp; Harpfer (2007); Stutzer &amp; Frey (2005 &amp; 2006); Weinzierl (2005)</td>
<td>Germany, GSOEP</td>
<td>Life satisfaction (0-10)</td>
<td>Life satisfaction (0-10)</td>
</tr>
<tr>
<td>Lelkes (2006)</td>
<td>Europe, ESS</td>
<td>Life satisfaction (1-10)</td>
<td>Life satisfaction (1-10)</td>
</tr>
</tbody>
</table>
### Table 16: Studies finding a negative impact of education on SWB

<table>
<thead>
<tr>
<th>Study</th>
<th>Dataset</th>
<th>Well-being question</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker et al. (2005), Thoits &amp; Hewitt (2001)</td>
<td>US, ACL</td>
<td>Life satisfaction (1-4)</td>
<td></td>
</tr>
<tr>
<td>Marks &amp; Fleming (1999)</td>
<td>Australia</td>
<td>SWB index with 9 items</td>
<td></td>
</tr>
<tr>
<td>Headey &amp; Wooden (2004)</td>
<td>Britain, BHPS</td>
<td>Life satisfaction (1-7)</td>
<td></td>
</tr>
<tr>
<td>Ferrer-i-Carbonell &amp; Gowdy (2005)</td>
<td>Britain, BHPS</td>
<td>Life satisfaction (1-7)</td>
<td></td>
</tr>
<tr>
<td>Helliwell &amp; Putnam (2004)</td>
<td>International, WVS</td>
<td>Life satisfaction (1-10)</td>
<td>Compared to not graduating high school, being a high school graduate significantly reduces life satisfaction, as does leaving between high school and university</td>
</tr>
<tr>
<td>Meier &amp; Stutzer (forthcoming)</td>
<td>Germany, GSOEP</td>
<td>Life satisfaction (0-10)</td>
<td>Significantly positive effect in OLS, becomes significantly negative in fixed effects</td>
</tr>
<tr>
<td>Headey &amp; Wooden (2004)</td>
<td>Australia, HILDA</td>
<td>Life satisfaction (0-10)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 17: Studies finding a non-significant impact of education on SWB

<table>
<thead>
<tr>
<th>Study</th>
<th>Dataset</th>
<th>Well-being question</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith (2003)</td>
<td>WVS, US and West Germany</td>
<td>General happiness (1-4)</td>
<td></td>
</tr>
<tr>
<td>van den Berg &amp; Ferrer-i-Carbonell (2005)</td>
<td>Netherlands</td>
<td>General happiness (1-5)</td>
<td></td>
</tr>
<tr>
<td>Dehejia et al. 2005</td>
<td>US, NSFH</td>
<td>General happiness (1-7)</td>
<td></td>
</tr>
<tr>
<td>Luttmer (2005)</td>
<td>US, NSFH</td>
<td>General happiness (1-7)</td>
<td></td>
</tr>
<tr>
<td>Helliwell &amp; Putnam (2004)</td>
<td>Canadian, ESC</td>
<td>Life satisfaction (1-10)</td>
<td></td>
</tr>
<tr>
<td>Flouri (2004)</td>
<td>Britain, BNCDS</td>
<td>How has life turned out so far</td>
<td>(0-10)</td>
</tr>
<tr>
<td>Ferrer-i-Carbonell (2005)</td>
<td>West Germany, GSEOP</td>
<td>Life satisfaction (0-10)</td>
<td>Some controls for individual heterogeneity</td>
</tr>
<tr>
<td>Oswald &amp; Powdthavee (forthcoming)</td>
<td>Britain, BHPS</td>
<td>Life satisfaction (1-7)</td>
<td>Those A levels and degree education report lower life satisfaction than those without qualifications in RE models but higher in FE models.</td>
</tr>
</tbody>
</table>
### Table 18: Studies showing the impact of education on mental health

<table>
<thead>
<tr>
<th>Study</th>
<th>Data source</th>
<th>Measure used</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magdol (2002); Kim &amp; McKenry (2002); Brown (2000)</td>
<td>US, NSFH</td>
<td>CES-D</td>
<td></td>
</tr>
<tr>
<td>Baker <em>et al.</em> (2005)</td>
<td>US, ACL</td>
<td>CES-D</td>
<td></td>
</tr>
<tr>
<td><strong>Not significant</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoits &amp; Hewitt (2001)</td>
<td>Britain, BHPS</td>
<td>GHQ</td>
<td></td>
</tr>
<tr>
<td>Lee <em>et al.</em> (2001)</td>
<td>US, NSFH</td>
<td>US (NSFH)</td>
<td></td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clark (2003) CHECK</td>
<td>Britain, BHPS</td>
<td>GHQ Caseness</td>
<td></td>
</tr>
<tr>
<td>Clark &amp; Oswald (1994)</td>
<td>Britain, BHPS</td>
<td>GHQ Caseness</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4: Criteria to judge whether a measure of well-being can be useful for public policy

4.1 Introduction

Chapter Two reviewed different ways of conceptualising and measuring well-being. Chapter Three highlighted the importance of the choice of measure and how well-being can vary considerably depending upon the measure. This chapter asks how we can choose between these conceptions and measurement instruments. It sets out a list of criteria to determine which measure will be most appropriate for public policy.

Setting out clear criteria will aid the process of making comparisons and highlighting limitations, both at a theoretical and empirical level. Empirical measures of well-being are easily criticised. However, where measurement is essential for policy, the perspective taken is how well a measure performs compared to current practice or the best alternative.

There are many areas of public policy in which a well-being measure could be used.

1. **Micro evaluation**, for use in: (a) local projects, where the change in well-being is not intended to be applied beyond the locality; (b) local projects, where it is assumed feasible to replicate the costs and benefits of the project elsewhere; and (c) interventions which impact on both quality and length of life (e.g. health care or safety interventions).

2. **Research into understanding well-being**, such as the causes of well-being throughout the life course, knowledge from which may indirectly contribute towards the direction of policy and government strategy.

3. **Macro policy design**, through provision of information to support fiscal and monetary policy.

4. **Monitoring equity concerns**, to develop and evaluate policy initiatives aimed at distributional issues.

5. **International comparisons** aimed at judging relative country performance, and contributing to national level policy agendas.

There will be different requirements upon a measure of well-being depending upon its desired purpose. However, it is useful to establish some general criteria of an ideal measure.
4.2 Choosing the assessment criteria

The criteria presented here were derived through considering existing criteria (for well-being, or measures related to well-being) and judging their merits from a public policy perspective, paying attention to whether they could be implemented.

The philosophical literature may have something to say about such criteria. The dialogue on well-being focuses mostly on establishing concepts which are robust to objections, rather than concepts which are suitable for policy. However, both Griffin and Sumner have proposed some recommendations on the conception of well-being required for moral judgments and policy, which form a useful starting point.

Griffin sets out three criteria “first, we need the account of well-being that we adopt … to be a plausible account of the domain of prudential value that it tries to cover; second, it must be what we want to use, for purposes of moral judgement, as the basis of comparison between different persons; and, third, it has to lend itself to the sorts of measurement that moral deliberation needs” (Griffin, 1986: 108).

Although aimed at moral judgment rather than resource allocation, these concerns can also be applied to resource allocation decisions. The last of Griffin’s criteria, that it lends itself to measurement, is clearly essential and will be discussed later under the general concern of empirical usefulness.

The first criterion assumes that the concept of interest is that of prudential value. This shall be adopted and is justified below.

The second criterion, that the conception is one that we would wish to use as the basis of comparison, emphasizes that the conception must be suitable for interpersonal comparisons. This can be thought of as both comparisons between people and across the same person over time. The need for interpersonal comparison arises both in terms of ensuring an appropriate conception and again in terms of the measurement instrument being interpersonally comparable, both issues shall be discussed below.

Griffin’s second criterion also specifies that the concept must be the one that we would want to use. This implies the need to make a normative judgement about the chosen conception of well-being. From the perspective of public policy, there may be components of well-being which would not be considered to be normatively relevant, and we may wish for a restricted conception of well-being to guide policy rather than a complete measure of well-being.
For example, it could be argued that a well-being measure for policy should exclude increases in well-being which are derived from sources considered to be illegal or socially illegitimate (see Hare, 1982). If the measure includes benefit arising from torturing animals, for example, then well-being would increase when the torture of animals increased, which seems inappropriate. However, would society consider the following changes in well-being as legitimate: an individual’s life satisfaction decreases when a gay couple move next door due to homophobic views; a young women positively assesses a moment when she believes men have looked at her appearance and found her attractive; a young women assesses taking illegal drugs, from a fully informed and non-addicted state, as positive? These examples show that the boundary of social legitimacy is likely to be difficult to define and will be socially determined.

Another contentious issue is when well-being is affected by comparisons to others. Of particular importance is whether to account for any reduced well-being following the increased income of another person? Harsanyi proposes that the social welfare function should exclude antisocial preferences, such as ‘sadism, envy, resentment, and malice” (Harsanyi, 1982: 56), suggesting that comparison income be excluded from impacting upon the measure of well-being. Similarly, Elster and Roemer (1991) argue that if the purpose of interpersonal comparison is to prepare the ground for transfer of resources, the ‘other directed’ components of well-being including sadistic and spiteful pleasures and altruistic pleasures should not count and “people should not lose claims on compensation if they happen to get a lot of pleasure from the pleasure of others” (Elster and Roemer, 1991: 6).

Whilst there may be some element of choice involved in deciding who to compare oneself with, other people’s consumption may also be viewed as imposing an unavoidable externality (Frank, 1997). Layard argues that desire for status is ubiquitous and hardwired into our biology (Layard, 2006). If preferences based on comparisons to others or desire to cause harm to others are excluded, should that extend to any preferences which cause harm to ourselves, such as greed, masochism, or attempted suicide?

The more measurement relies upon subjective assessment the more at risk it will have of including well-being from sources which may conflict with widely held social values, and the further it may deviate from a level of well-being which in societies view the individual ‘ought’ to experience, or that level which society may wish to take into consideration in resource allocation decisions. However, aside from practical issues of actually operating such restrictions, defining when a change in well-being should or should not be included, as the above examples demonstrate, will not be clear cut. If and when well-being should be stripped of its anti-social element is a judgement which is tied to current conventions and social norms. Incorporating these normative judgements within the measurement process runs the risk of them
not being exposed to dialogue and debate. It is therefore preferable to include a *prima facie* measure of well-being from the individual’s perspective. This will help guard against data being excluded on ideological grounds.

An argument could be made for restricting the concept of well-being to only those parts for which society can be held responsible, and excluding parts which are the responsibility of the individual (Roemer, 2002). For example, we may think that benefits arising from individual effort (such as undertaking regular exercise) should not be included. However, where this boundary lies, between an individual and society, is subject to debate.

Similarly, well-being may arise from areas which are not perceived as appropriate concerns for government intervention, such as religious beliefs. However, attributes of well-being should only be excluded from a policy measure if they could not not be impacted upon by any government policies (something which is unlikely to be known in advance). Whilst it may be appropriate to exclude some attributes of well-being from being targets of policy, the consequences of any action designed to improve well-being should be judged by its impact on well-being overall. For example, if policy to improve satisfaction at work (which could be argued to be within government remit) was judged purely in terms of changes to satisfaction at work, this may overlook indirect consequences the policy may have on satisfaction with family life (which could be argued not to fall within the remit of government). Consequently, this type of restriction at the measurement level is not useful. An ideal measure would include all consequences on an individual’s well-being, whether these are considered an appropriate remit of policy is a separate question.

Lessons may be learnt from the restriction of utility in welfare economics to ‘economic welfare’. Pigou saw this as a practical limitation (Pigou, 1932). However, ‘economic welfare’ became synonymous with ‘welfare’, and other aspects of welfare, because they were not included within the measurement, were overlooked. Restricting measurement to aspects that are currently perceived as the appropriate concern for government risks ignoring the impact of policy on other aspects of well-being. A more constructive approach is to make the measurement of well-being inclusive, complete and separate from political considerations.

In addition, there may be attributes which are part of well-being but which the government is unable to do anything about. For example, well-being may be affected by unchangeable personality traits. If the well-being measure excludes the contribution of a permanent trait this will not be a problem for achieving an efficient allocation of resources, however, a complete measure of individual well-being will still be required for equity considerations.
Consequently, although we may have reasons to ‘want’ a particular account for moral comparisons, the criteria of normative morality is not included here. This is in part due to a desire to separate out moral issues from well-being measurement, and in part because it is unlikely to be attainable from a practical measurement perspective.

Turning to Sumner’s criteria, he argues that theories of welfare should have ‘descriptive adequacy’, which includes completeness (allowing judgments of individual welfare relative to previous levels and other people), formality (focusing on the nature of well-being rather than its causes), neutrality (free of distorting biases or favouring particular forms of life), and fidelity (faithful to our ordinary understanding of well-being and compatible with our widely held intuitions about well-being) (Sumner, 1996).

From a policy perspective, completeness and fidelity will be essential. However, his other criteria (formality and neutrality) are less useful. At the theoretical level a distinction between input and output is essential for clarity of the concept, yet at the measurement level some degree of overlap may enhance accuracy of the measurement instrument hence we may not wish to rule out instruments because they capture some aspect of input rather than output. This would violate Sumner’s formality criterion, as the inputs are the causes of well-being. For example, measures that focus on psychological needs are essentially input based, yet may offer a good proxy of well-being.

The neutrality condition pre-supposes a particular conception well-being (Toner, 2006), and whilst avoiding bias is clearly desirable, we do not wish for the criteria to pre-judge available options and consequently require a less restrictive notion of bias. Government policy unavoidably makes normative judgments. The presence of a normative judgment within the concept of well-being does not therefore exclude it from being a useful measure for policy.

Our intuitions about well-being involve current assessment and predicting into the future. Given our inability to accurately predict future outcomes (discussed in Chapter One) it seems unlikely that all our intuitions about how well our lives will go are necessarily correct. Even our reflective judgments may mislead us, suggesting a degree of judgment may be needed about which intuitions we would wish to honour in choosing the conception of well-being. This illustrates that care is needed with regard to Sumner’s criterion of fidelity. Whilst being faithful to ordinary understandings about well-being seems highly reasonable we require criteria which are sufficiently flexible to deal with situations where our intuitions about well-being are incorrect.
In summary, the concept of well-being should be (i) stated only in terms of prudential value, (ii) interpersonally comparable, (iii) complete, and (iv) measurable. We can turn to the social indicators movement, sociology and psychology literature to consider possible means to judge completeness and measurability.

The International Society for Quality of Life Studies oversaw a committee that developed 14 criteria for reviewing quality of life (QoL) indexes for public policy (Hagerty et al. 2001). This exercise began from the perspective of requiring a domain-weighted index in which the domains are measured both objectively and subjectively, and their criteria reflect this prior choice. Since the aim here is to compare not just different subjective measures but different approaches to thinking about well-being, the restrictions implicit in Hagerty et al. (2001) make these criteria unsuitable.\(^{43}\)

Methods for judging the quality and appropriateness of measurement instruments are well developed in psychology, and have been applied to QoL measures (Fitzpatrick et al., 1992), health-related quality of life (HRQoL) measures (Anderson et al., 1993; Sturgis et al., 2001) and SWB measures (Diener, 1994). The psychometric literature uses a range of criteria for assessing the performance of instruments, and established empirical methods. These include practicality (can the instrument be easily used for the population and purpose intended?), internal consistency (do the different items which are intended to cover a single dimension correlate sufficiently highly to suggest they are tapping into a single dimension?), reliability (can results be reproduced over repeated administrations, between different interviewers or raters and between different times and places of administration?), responsiveness (is the instrument responsive to change in the construct being measured?), and validity (does the instrument measure what it purports to measure?). These tests capture two general concerns, firstly whether the measure is valid and secondly whether it is practical and fit for purpose.

Such psychometric tests, whilst extremely useful, have not been seen as sufficient for assessing measures to be used in economic evaluation (Brazier and Devrill, 1999), in part because the end point for psychometric tests (health measures in the case of Brazier and Devrill, 1999) is often only an input into preference based measures. Brazier and Devrill note that measures of internal consistency, for example, could conflict with the requirements for economic evaluation since “it may result in the exclusion of items that do not fit neatly into one of the hypothesized dimensions but are important in terms of patient or societal preferences” (Brazier and Devrill, 1999: 42). This highlights a danger that an instrument may appear successful across a range of psychometric criteria, yet be invalid in part due to an absence of theoretical foundations.

\(^{43}\) These criteria present a useful way of thinking about redundancy of information. This is adopted later.
Economists have traditionally emphasised whether instruments are drawn from a consistent theoretical approach, namely whether instruments adequately reflect individual preferences, and given less emphasis to psychometric testing of validity or to the underlying validity of preference satisfaction as a measure of well-being.

A desirable measure of well-being for policy will be one which is valid at many levels, from descriptive adequacy of our theory of well-being to the measurement validity of the chosen instrument. For example, if the measure is expenditure diaries then it is necessary to consider whether these are a valid measure of consumption, whether consumption is a valid measure of preference satisfaction, and whether preference satisfaction is a valid measure of well-being. Traditional psychometric tests will inform about the first level, but high validity at this level does not guarantee a good measure of well-being.

Without a gold standard for well-being or prior commitment to a particular account of well-being, validity can only be considered indirectly. Validity will be considered here by exploring whether the measure has: i) content validity and covers all the important aspects of well-being and is constructed from a process which ensures completeness; ii) convergent validity and is positively correlated with those attributes which we think of as being related to high well-being; and iii) predictive validity and accurately predicts behaviour or future states hypothesized to be linked to changes in well-being.

In addition to validity, the measure will need to be fit for purpose. Whether it is empirically useful is likely to be less controversial, however, these concerns may be overlooked by those who do not have policy objectives as a primary concern.

For a measure to be empirically useful for policy it must be responsive to changes in well-being, efficient and practical to collect and have low and unbiased measurement error. Usefulness of the measure, and take-up of regular monitoring, will also be influenced by whether it is perceived as legitimate by the public and policy makers.

The measure will be most useful for policy if it can summarise well-being within a single index, and even more so if that index can appropriately be treated as cardinal. The advantages of cardinality are greater in government policy than traditional uses of psychological tests, and is therefore emphasised. Whilst psychometric testing often includes consideration of ceiling and floor effects, cardinality of measures is often assumed without theoretical or empirical support.

In summary, well-being measures will be assessed against 10 criteria. Two that help distinguish whether the measure is referring to the appropriate concept of well-being: that it is a measure of prudential value only and that it is interpersonally comparable. Three that help confirm the
validity of the measure: that it has content, convergent and predictive validity. And five which help establish whether the measure will be empirically useful: that the instrument has low and unbiased measurement error, is cardinal, sensitive, practical and efficient to use and compatible with public views. For each of these criteria it is necessary to consider the measure at both a theoretical and a practical level. The remainder of this chapter will further explain and justify the importance these criteria.

4.2 Appropriate conception of well-being for public policy

4.2.1 Prudential value only

In Chapter Two a distinction was made between the ‘good life’ and a life which is good for the individual. The concept of a ‘good life’ includes values beyond prudential value such as moral, spiritual, and aesthetic concerns. Some theories of well-being are only concerned with whether life is good for the individual, others contain other values. Some of the differences between a concept of well-being as the ‘good life’ and as good for the individual arise because of different perspectives from which assessment can be made (the individual’s perspective, society’s, or an omnipresent being’s) and from different time periods from which the assessment can be made (during the individual’s lifetime, at the end of their lifetime, 100 years after their death etc.).

Whether something that the individual does not experience or even know about can affect their well-being is a question which has generated much philosophical debate. For example, if an individual is deceived but never finds out, does this affect their well-being? Or if an individual achieves success once they are dead, does this affect their well-being? Whether or not these are relevant to an individual’s well-being, they would be difficult to incorporate into policy for the following reasons. Pragmatic measurement issues mean their inclusion would be impossible (consequences after death may be virtually infinite), but equally important, public agreement on their importance to well-being is unlikely. The perspective best suited to government policy is the individual’s, during their lifetime.

Justifying government action based on something being ‘good’ but not ‘good for the individual’, or the citizens the government represents, is clearly problematic. That is not to say that non-prudential values should not be concern for government. However, for transparency and clarity, and because of measurement problems, these should be separated out from well-being. A similar stance is taken by Adler and Posner (2006), who present deontological concerns as separate, additional concerns for policy makers. Consequently, the well-being measure sought here is a measure of prudential value only.
Chapter 4  Criteria to judge whether a measure of well-being is useful for policy?

4.2.2 Inter-temporally and interpersonally comparable

A concept of well-being which is interpersonally comparable does not deny that some qualities of people’s lives may be non-commensurable. For example, we may not be able to compare the ‘worthiness’ of two people in terms of comparisons upon moral and spiritual dimensions. However, the comparison we are interested in here is restricted to a comparison of prudential value.

Economists have been reluctant to engage in interpersonal comparisons of utility, and expressed far more concern in this regard than psychologists. For example, Robbins held that, “Introspection does not enable A to measure what is going on in B’s mind, nor B to measure what is going on in A’s” (Robbins, 1932 [1945]: 140) and any assumption of equal capacity of satisfaction is seen as a normative, ethical judgment without possible empirical verification.

In the absence of being able to make interpersonal comparisons of well-being, the choices over policy options can turn only to considering the inter-temporal changes in well-being. If changes in well-being across different people cannot be compared then a policy option can only be taken as an improvement when the change is preferred by at least one individual and all others are either indifferent or also prefer the change. However, as discussed in Chapter Two, the Pareto condition provides no guidance on matters relating to changes in the distribution of resources, and no guidance on how to choose between alternative Pareto optima. We cannot escape the fact that much collective decision-making involves, comparisons of the costs and benefits incurred by different people. If governments are to take well-being into consideration for policy they need to be able to compare how well life is going for different people. As stated by Hausman,

“...if a conception of well-being does not permit one to make interpersonal comparisons in an acceptable way, then that conception of well-being is itself unacceptable” (Hausman, 1995: 474).

Interpersonal comparisons have been argued to be reasonable (e.g. Ng, 1997). The validity of interpersonal comparisons is often based on imaginative empathy and an “admission that other people are just as real as we are … and that they have the same basic capacity for satisfaction and for dissatisfaction, in spite of the undeniable individual differences that exist between us in specific detail” (Harsanyi, 1982: 52).

Others have argued that even if we think that individuals are similar to ourselves this does not imply measurement of utility that is interpersonally comparable.

“Let us assume that there is a fact of the matter in an interpersonal comparison of well-being, considered as hedonic enjoyment or satisfaction of desires. It does not follow that
we could ever discover it … other minds are just as inaccessible to us as the past. We need not entertain doubts about their existence and their essential similarity to our own, but we may despair at ever getting the details right” (Elster and Roemer, 1991: 11).

Consequently, requiring interpersonal comparisons is likely to be less problematic at a conceptual level than a measurement level. However, some margin of error will be acceptable, hence the essential criteria is for rough interpersonal comparability.

Measurement error will arise due to problems measuring individual well-being. Further measurement problems arise when we try and make comparisons. However, comparisons between subgroups of people or between the same subgroup at different periods of time should still provide useful inputs into policy. As long as there is no reason to suppose systematic variance in use of the instrument as a function of group membership any differences between groups are likely to be meaningful, irrespective of small variations in the use of the instrument within groups. However, we may hypothesise a range of reasons why instrument use may vary with group membership, for example, those with good health may interpret the top of the life satisfaction scale differently to those with poor health.

Comparing measurement of well-being of the same individual at different points in time will be subject to inaccuracies if factors relevant to the measurement process change over time. For example, over time the individual may change their preferences, beliefs and modes of judgement.

Interpersonal comparability requires that measurement scales referring to different people can be meaningfully compared, and that person A’s 7/10 represents the same level as person B’s 7/10. It will be difficult to prove whether this is the case. However, it can be explored indirectly by considering whether the use of the measurement instrument appears to differ between people. We can consider whether members of society understand language used in the question in the same way, and whether they make similar judgements about the well-being of other individuals.

4.3 Validity

4.3.1 Content validity

In the case of well-being, content validity assesses whether the measure incorporates all attributes that are seen to make someone’s life better, hence is an exhaustive measure of well-being. For example, if the measure excludes something an individual could have and society cares about whether she has it for her sake, then the measure may not be sufficiently complete. If an attribute which we care about people having is not picked up by the measure, and this attribute differs between members of the society, then either we need to cast doubt on our
intuitions that the attribute is something we should care about or cast doubt on the measure of well-being.

Content validity incorporates ‘face validity’ or whether the measurement looks like it measures what it intends to and is compatible with widely held intuitions about well-being. The measure should ‘on the face of it’ appear to be a sensible and appropriate.

### 4.3.2 Convergent validity

Convergent validity assesses whether the measure converges with other measures that should be theoretically similar. If the measure is not correlated with factors which we take to be signs of individual well-being, such as health, material resources, facial expressions, opinions of the individual about how their live is going and opinions of those close to the individual, then this would be of concern for the measure unless a reasonable explanation could be offered. Since our intuition on sources of well-being may be incorrect, correlations of a measure of well-being with an attribute that is commonly thought of as being a source of well-being should be treated with caution.

If one measure of well-being offers a different magnitude of well-being to the other measures, then this gives us reason to doubt the validity of the outlying measure.

### 4.3.3 Predictive validity

A measure has predictive validity if it can correctly predict something we theoretically think it should be able to predict. For example, low well-being should predict attempts at suicide. Where the measure is composed of sub-components predictive validity may be tested based on predictions arising from these sub-components. For example, job-satisfaction may be tested by predictions of quitting, marriage-satisfaction by separation, and area-satisfaction by moving house.

### 4.4 Empirically useful

#### 4.4.1 Reliable with low and unbiased measurement error

The measure should not include irrelevant attributes. These may arise at a theoretical level (for example, the inclusion of non-prudential values), or at the measurement level (for example,
framing effects within survey questions). The measure of well-being should change only when well-being changes\(^{44}\), hence test-retest reliability should be high in the absence of such change.

If measurement error is purely random, taking the average of a number of responses should result in a more accurate measure. The measure will be used to support decision making using averages of groups rather than individuals, therefore problems which relate to only few cases and random measurement error will not be excessively prohibitive in the measures usefulness. As noted above, problems arise when measurement error is related to group characteristics, in which case group level measurement will not overcome the existence of measurement error.

### 4.4.2 Cardinality

Measurement requires that the concept be numerically represented. The measure must be at the very least be ordinal for it to offer any guidance to policy such that a higher level of the measurement instrument must indicate higher well-being. This requires a single index of well-being.

Where policy makers are faced with more than one attribute to represent well-being, trade-offs between attributes will be done at an implicit level within policy decisions. More useful for policy makers is a single index in which different attributes are already assigned weights to indicate their importance to well-being.

Rank orderings of well-being are of restricted use for policy unless the magnitude of the differences is also known, and well-being levels can be aggregated and averaged across groups, hence the need for a cardinal measure. Requiring a cardinal measure assumes that it is meaningful to treat the both underlying concept and the measurement instrument as cardinal.

Since Pareto (1906) it has been widely accepted amongst economists that utility is merely a representation of preference orderings, and that cardinal utility is unmeasurable from choice behaviour revealed through market purchases (Robbins, 1932; Hicks and Allen, 1934; Samuelson, 1938). Economists adopting an ordinal preference satisfaction framework have been reluctant to think in terms of getting a certain times as much satisfaction from x (say going to the cinema) compared to y (say watching a DVD) leaving comparisons as rank orders only. Furthermore, such problematic assumptions were not necessary for consumer theory. However, because ordinal utility is not required by consumer theory and not revealed by market decisions, does not imply cardinality is impossible. Some economists have continued to argue in favour of

\(^{44}\) Although there may be difficulties determining what well-being is, this criterion is useful as we can tell when it is measuring something that we know is not well-being (e.g. the weather at time of measuring).
cardinal utility (e.g. Frisch, 1964; Ng, 1997). Ng argues that welfare is fully cardinal and people can make statements about how much more or less happy they are in one state over another (Ng, 1997: 5). However, Ng also admits that it is not possible to make precise comparisons, but he asserts that this imprecision is due to imperfect memory and is little different from asking an individual to compare different amounts of water they have drunk.

At the individual level, we do not seem to find it difficult to make decisions between ways of spending our time, money and energy on things that yield non-comparable benefits: it seems an individual can get by without having a clear idea of exactly how much additional well-being she gets from owning a cat, her relationship with her sister, eating chocolate ice cream or having a detached garage. Griffin describes our individual judgements of intrinsic reward as ‘roughly cardinal’ and capable of distinguishing big from small differences finely enough to guide individual decisions on what is worth sacrificing, or risking, now for future gains (Griffin, 1986).

We ideally want governments to have a clear idea of the benefits to its citizens from expenditure it makes on health care, education etc., and a cardinal measure of well-being would enable them to compare marginal costs with marginal benefits and move towards an efficient allocation of resources. Without a cardinal measure of well-being trade-offs will still be made between benefits in different areas, hence benefits are will be implicitly treated as at least roughly cardinal. The use of a cardinal measure of well-being should enable these trade-offs to be more explicit.

4.4.3 Appropriate sensitivity

A reliable measure (4.4.1) will by definition be sensitive to changes in well-being. Appropriate sensitivity requires that important and relevant differences in well-being can be detected. From a policy makers perspective being able to show changes that arise due to altered circumstances over which the government has some degree of control is essential.

Measures should not be so sensitive that they detect differences that are not relevant to the individual or to policy. Increases in apparent sensitivity may be made at the expense of a reduction in validity. A parallel situation may be found in health care evaluation where a choice may be made between using more sensitive disease-specific HRQoL measures or generic

---

45 In other areas of economics the concept of cardinal utility remained. In decisions under uncertainty, Neumann-Morgenstern’s expected utility hypothesis offers a means of measuring cardinal utilities by constructing an expected utility function from the utilities of possible outcomes weighted by their probability.
HRQoL measures, which are potentially more valid as an overall measure of HRQoL but at the expense of a loss in sensitivity.

The degree of sensitivity required depends on the context to which the well-being measure is to be applied. If it is to be used for national level comparisons of well-being – either over time or between large groups – it may be possible to detect significant differences with a fairly insensitive measure. However, an outcome measure for smaller scale policy evaluation requires a greater degree of sensitivity. Ideally, the measure should detect any change in individual well-being which is politically relevant, either for a judgement of efficiency or equity. It is more important to detect substantive significant differences than merely statistically significant ones (McCloskey and Ziliak, 1996), however, we have little feel for what makes an important change in well-being, either at the individual or policy level.

Sensitivity should be consistent across the full range of the measure, such that the instrument can distinguish between changes in well-being when it is high or low. Clustering at one end of scale may imply lack of ability of the instrument to discriminate at that end of the scale. Ceiling effects, where the measure fails to distinguish differences at very high levels) and floor effects (failure to distinguish differences at low levels) may suggest that cardinality was not holding. However, in both cases this may also be a reflection of true distribution of well-being.

4.4.4 Practical and efficient to collect

For purely practical reasons any measure of well-being for public policy must be easily and cheaply attainable. Cost and response rates are likely to be determined by completion time and mode of administration and the perceived sensitivity of the information.

In order to measure well-being efficiently there should not be redundancy within the instrument, additional items should only be included if they provide additional information. The criteria developed by Hagerty et al. (2001) give restrictions on additional domains based on their unique contribution to the overall index of well-being. This approach can be applied to measures other than domain-weighted satisfaction, such as those drawn from substantive good accounts.

4.4.5 Public/political acceptability

A measure of well-being for public policy will fail to be useful if it is met with resistance from the public to incorporate findings from the measure into policy making.

If the measure of well-being incorporates autonomy and public preferences, such that well-being can only increase if public wishes are met, a Pareto improvement can not contravene
public wishes. Preference satisfaction is widely supported within economics in part because of consumer sovereignty and the idea that the individual knows what is best for themselves.

If measurement does not incorporate individual preferences, well-being could (at least in principle) increase from an action which the public would not support (such as enforced restrictions on consumption of junk food\footnote{Gruber and Mullainathan (2005) provide an interesting study in which the increased taxation upon cigarettes results in greater reported happiness for predicted smokers.}). Individuals may even agree that such actions would enhance their future well-being yet still not support them if they placed greater value on their autonomy and freedom and may choose lower future well-being rather than sacrifice this freedom. This raises concerns that policies to enhance well-being may not have public support and may be perceived as a threat to individual sovereignty. McMahon (2007) for example, raises fears about “reducing individual sovereignty to self-reports of one’s well-being.”

However, incorporating well-being into policy making does not necessarily imply that well-being is the only outcome of concern. It is possible for the government to know that $x$ will increase an individual’s well-being without that leading to any policy to increase $x$, particularly if to do so would be against the public’s wishes. The freedom and autonomy of citizens may indeed be something policy makers wish to respect. If these attributes are purely instrumental to well-being there will be no conflict. If they hold additional values policy may legitimately trade-off increases in well-being to protect or enhance these other values. Measures that do not incorporate autonomy do not imply a disregard for autonomy. Incorporating autonomy within the measure conflates two separate issues, and is not a necessary criterion for a measure of well-being for public policy. Griffin notes, “it just confuses two quite different ideas to adopt the actual desire account of well-being just because it makes autonomy prominent” (Griffin, 1986: 11). Maximising a single welfare function of prudential value would not be a sensible objective where other values (truth, justice, freedoms, knowledge, beauty etc.) were regarded as important.

If the measure is not perceived as relevant to individual lives or to policy, either directly or indirectly, its successful implementation will be in doubt. However, public views are not static. One account or measure of well-being could be considered by the public to be most suitable at one point in time but this could change due to moral debate, political discourse, or changes in public preferences. If public views have been influenced by the existing measures used in policy, then they may well change with time if alternative measures are used. Whilst it would be naïve to ignore public and policy maker acceptability this should not override validity concerns.
4.5 Conclusion

The appropriate conception of well-being, from a public policy perspective, is prudential value that can be measured, and is interpersonally comparable. In order to maintain transparency, and measurability, non-prudential values are not to be included in the concept.

Of all the criteria validity is the most important. A measure that appears ideal in all other respects but is not actually measuring well-being will not be fit for purpose. Compromises may need to be made in terms of practicality, however, the valid conception of well-being should not be lost.

Concerns over data availability have not been raised. This is because the importance of knowledge of well-being justifies any additional data collection. If data is not currently available, it may become available in the future (see Robeyns, 2005).

Judging the validity of well-being measures is problematic since there is no gold standard with which to compare the measures. However, the various criteria aimed at assessing validity (content, convergent and predictive) should enable a good judgement to be made on whether the measure can be successfully defended.

The measure also needs to meet certain criteria relating to practical usefulness. These include reliability and low, un biased measurement error, cardinality, sensitivity, practicality, and having political legitimacy. These are all considerations that will influence how useful the measure could be in practice to policy makers. The importance of each of these criteria will vary depending on the purpose to which the instrument is being used.

If the measure is required for small-sale project evaluation that is not intended to be applied beyond the location (1a) it may be possible for participants of the project to be involved within the establishment of the measure, and perceived legitimacy may depend upon such involvement. However, extensive participation will become impractical for larger scale projects or programme evaluation. Where the purpose of well-being measurement is to evaluate locally specific projects (1a) issues of comparability outside of the locality are less of a concern. However, where comparisons are required between different people (1b) a more universally comparable measure of well-being is necessary. Where a measure is sought for international comparison (5) interpersonal comparability requires comparison across culture and nationality. Although these issues will also arise to some extent within multicultural country, such as the UK, within country cultural differences would be expected to be less extreme than those between countries.
Consequently, the relative importance of these criteria will vary depending upon the purpose of the measurement instrument. It is unlikely that any well-being measure will fully meet all criteria, suggesting an additional judgement will be required. There may also be trade-offs between the various criteria, for example, measures high on test-retest reliability may perform poorly in terms of sensitivity. Practicality may require a trade-off against validity and/or sensitivity. The value of having such criteria is transparency — we can see what trade-offs are being made.

Given the ongoing theoretical dialogue surrounding well-being, and the measurement error present in most individual characteristics (income for example), a strict criteria of validity and accuracy in measuring well-being is unlikely to be achievable. The important consideration is whether concerns over accuracy are such that the measure is unlikely to be useful for resource allocation decisions.

The next chapter assesses different accounts and measures of well-being within different accounts according to whether they meet the criteria identified here.
Chapter 5: How closely do well-being measures meet the ideal criteria?

5.1 Introduction

This chapter uses the criteria set out in Chapter Four to assess theories of, and measures of well-being, including preference satisfaction, focusing on income and consumption and health related quality of life (HRQoL) measures, and subjective accounts including affect based measures and evaluative life satisfaction measures.\(^47\)

5.2 Appropriate conception of well-being for public policy

In Chapter Four I argued that well-being (for public policy) should be conceived as (i) prudential value, (ii) interpersonally comparable, and (iii) susceptible to measurement.

5.2.1 Prudential value only

Preference satisfaction

Preference satisfaction can incorporate prudential and non-prudential values. An individual may prefer a particular life not because they see it as better for themselves but because of other attributes associated with that life, such as the contribution it brings to others.

Of course, what appears to be non-prudential concerns may also be sources of prudential value. For example, actions motivated by helping others may simultaneously satisfy underlying prudential desires, such as the desire for a positive self-image. However, actions which dramatically reduce an individual’s life, such as jumping in front of a bus to save a child, are impossible to rationalise unless we allow that individuals can be motivated by more than a desire to maximise their experienced well-being. If an individual willingly chooses to make a sacrifice for another individual, it would seem appropriate from a public policy perspective for their post-sacrifice well-being level to be lower than the pre-sacrifice level, regardless of the fact that they had their wishes met in following out the sacrifice. The difficulty rests with equating maximising preference satisfaction with choice, rather than relying on preference satisfaction, since the individual may be better off for having the preference to save the child met, but worse off because underlying preferences they held to towards their own safety and their own future goals not met. That act of choice does not

\(^47\) For completeness, substantive good measures are given the same treatment in Appendix III.
necessarily imply that the former desire is greater than the latter. Darwall notes that it does not follow that if something is for someone’s good that there is any normative reason for him to desire it, since he may be indifferent to his welfare, or hate himself (Darwall, 2003).

An individual may care about and have desires for non-prudential values, such as the well-being of another species. If consumption is used to proxy for desire satisfaction (such as a donation to protect penguins), then allocating resources towards those ends assumes that the individual’s well-being has increased in proportionately. This is not then measuring prudential value.

Although in theory the inclusion of non-prudential values presents a problem for the use of preference satisfaction as a measure of well-being for policy, the proportion of our non-prudential desires is likely to be limited. The National Council for Voluntary Organisations estimated that the total amount donated to charity in 2004/05 was £8.2 billion in the UK, equivalent to an average monthly donation of £14.17 per head. Whilst many non-prudential desires may not reveal themselves in financial giving, the level of expenditure involved in charity donations suggests that expenditure on non-prudential concerns is not large. Furthermore, this type of consumption may have prudential benefits to the individual, such as connecting them with others.

Although the problem should not therefore be overstated, the inclusion on non-prudential concerns within preference satisfaction does create a lack of transparency within the measures of well-being based upon this account. An area where non-prudential concerns are problematic is the use of WTP for non-market goods that have aspects of bequest or existence value (Krutilla, 1967). Migrom (1993) showed that altruistic concerns can lead to double (or more) counting of benefits. The correct role of non-use values within cost benefit analysis remains unclear (see Johansson, 1992; Quiggin 1998). What is clear is that a failure to question the source of individual preferences creates uncertainty on exactly what is being valued.

In terms of HRQoL measures, preferences between health states and risk of death or reduced length of life may incorporate consideration of the impact a particular health state or risk of death has upon others, particularly potential carers and dependents. Whilst valuations may have aspects of non-prudential value it is very difficult to extract the prudential and non-prudential influences upon individual preferences towards different health states.

*Subjective: Hedonist*

Well-being as defined by hedonism, and measures compatible with this account, only incorporate how well a life is going for an individual and therefore captures only prudential value.

Subjective: Evaluative

Evaluative measures have the potential to incorporate non-prudential values. Life satisfaction responses may include an individual’s assessment of the lives of those close to them, or the world more generally. Life satisfaction may even arise from sacrifices to an individual’s own interests, such as sacrifices for their children. Cross-cultural studies have found that less individualistic cultures may incorporate concerns for family and community within their own life assessments (Lu and Gilmour, 2004). Some evaluative style survey questions may have more potential to incorporate non-prudential values than others, for example, questions focusing on living the ‘best life possible’ may lead individuals to think about moral or spiritual values in addition to prudential values.

In some cases the incorporation of what may look like non-prudential concerns (such as trust, creativity, and justice) may be indirectly incorporating concerns with long-term well-being. In others, apparent sacrifices for others may enhance prudential well-being. Whether an individual judges that she is leading a ‘good life’ and how she would judge her life in terms of how well it is going for her, are likely to, in most cases, be extremely similar.

Summary of providing a measure of prudential value only

Hedonist theories are the cleanest in terms of measuring only prudential value. Under WTP something can be counted as well-being when it is clearly non-prudential. This is of particular concern with the inclusion of existence values within WTP for non-market goods such as remote wilderness or species of animal. Although non-prudential values may be incorporated into an individual’s assessment of their life, the self-referencing nature of the SWB measures make this less problematic than for WTP.

5.2.2 Inter-temporally and interpersonally comparable

Preference satisfaction

The utility derived from preference orderings says nothing about how one individual’s utility compares with another individual’s utility, hence the reliance within welfare economics upon
the Pareto condition, or extensions to it (Kaldor, 1939; Hicks, 1939), to make judgements about changes in social welfare.

Chapter Two discussed the conceptual approach of seeing well-being as a position within an ordinal preference ranking of possible lives. Thinking of each life as holding a unique position within a preference ordering opens the possibility of interpersonal comparison within a preference satisfaction framework. Making interpersonal comparisons between the utility levels of two individuals involves asking how much satisfaction each of them derives from his own objective position, given his own preferences and personal attitudes (Harsanyi, 1996). This comparison could be conducted from a third-person perspective, using psychological laws to understand what life would be like for these two individuals. However, Harsanyi notes that this is inadequate due to our insufficient understanding of psychological laws. A preferred method is to take a first person perspective and try and achieve ‘empathetic understanding’ of what it is like to be in either individuals objective positive with their own attitudes.

“When different people make interpersonal comparisons of utility between the same two individuals or between the same two social groups, they may perhaps arrive at somewhat different conclusions, but most of the time their conclusions will be close enough to show that making such comparisons is not an altogether hopeless undertaking.” (Harsanyi, 1996: 144)

As discussed in Chapter Two, health state utility valuations adopt a similar process in which individuals consider what life would be like for someone experiencing a particular health state. However, the individual usually adopts her own preferences and attitudes for the valuation. Public preferences have been found to be significantly different to those of patients (Gabriel et al., 1999; Sackett and Torrance, 1978). Whose preferences should be used where preferences differ, is a matter of contention (Dolan, 1999).

From a perspective of aggregating satisfied desires, interpersonal comparisons require a judgement to be made on the extent of desires satisfied across different people. However, interpersonal comparisons vary depending upon how desires are treated and whether well-being is taken as the summation of met desires or incorporates both met and unmet desires. Take two otherwise identical people, A and B, who have identical met desires except that B has an additional desire, say for a Gucci handbag. If the desire for the handbag is not met, then based on the summation of desires met they have the same well-being, but based on percentage of desires met B has lower well-being. If the desire for the handbag is met, based on the summation of desires A has higher well-being, and based on percentage of desires A will have slightly higher well-being.
Income as a measure of preference satisfaction assumes that only those desires that are met contribute to well-being, or that unmet desires will be constant, or at least independent of income.

A serious obstacle to interpersonal comparisons of desire satisfaction arises where desires are influenced by existing and past circumstances. If desires are bought in line with expectations, then in situations of long-standing deprivation, for example, people may be “too subdued or broken to have the courage to desire much” (Sen, 1992). As explained by Griffin (1986), “Our desires are shaped by our expectations, which are shaped by our circumstances. Any injustice in the last infects the first” (Griffin, 1986: 47).

Similarly, Jon Elster argues that people will “tend to adjust their aspirations to their possibilities” (Elster, 1982: 219). Adaptive preference formation arises due to habituation and resignation, and preferences are therefore reversible and depend upon recent past experience. Since the individual’s preferences will be different in different periods of time depending upon recent experience, the order of preference for different circumstances will break down suggesting inter-temporal comparisons within one person become impossible.

Since everyone’s preferences in some sense will respond to circumstances, and will be conditional upon their own expectations, it does beg the question of just how similar affective response and assessment would need to be before we could be confident that preferences were not being infected by, say, injustice. Comparisons may require the assumption of autonomous preferences or those we would not mind having even if we knew how they had been formed (Elster and Roemer, 1991).

The only measures available that proxy for the aggregation of preferences are income or consumption. Comparing measures of income face two general problems. Firstly, income requires adjustment to account for diminishing returns to income before a comparison can be made. Difficulties with estimating the elasticity of marginal returns to income were discussed in Chapter One. Without constant marginal returns to income (or an appropriately adjusted income measure), such that individual desires can be equally met across people by income (or adjusted income), it will not be possible to make interpersonal comparisons of income as a measure of desire satisfaction.

Secondly, the measure we require is real consumption, accounting for both prices and decisions made in which either higher prices or reduced consumption reflect non-observed benefits. Comparing income between or within countries must account for differences in price levels. However, we may not wish to fully account for the differences in prices between living
say in London versus Northern Scotland since the higher living costs in London may be compensated for in other ways, particularly where there is flexibility in labour and housing markets.

*Subjective: Hedonist and Evaluative*

Interpersonal comparisons of subjective measures of well-being face both conceptual and measurement problems. At a conceptual level, feelings and evaluations may be considered an inappropriate indicator to make interpersonal well-being comparisons where there is reason to think the assessments are impaired.

This may arise if the individual was incapable of making a reasonable judgement about their feelings or how satisfied they are with their life due to mental impairment or unreliable judgement. Although at a theoretical level mental impairment does not undermine a hedonist account of well-being in the same way, since individuals still experience pleasure and pain, although finding this out may be problematic.

Subjective measures of well-being face similar problems to desire satisfaction accounts in terms of adaptation to poor life circumstances, satisfaction possibly arising from low expectations and the impact of past experiences. The “happy slave” phenomenon clearly demonstrates the problem. The fact that subjective assessments of satisfaction are partly determined by levels of aspiration has been seen as justification against using subjective measures of well-being for societal planning (Eriksen, 1993; Lerner, 1997).

This problem undermines both interpersonal and inter-temporal comparisons. Individuals may adapt to their circumstances to the extent that they feel positive affect and evaluate their lives as good but would give a different evaluation were they to gain different knowledge or experience. An individual’s assessment and their felt experience seems a poor indication of well-being where it has arisen from a position of ignorance and preferences which have been manipulated through the previous existence of certain states. The evaluative account could be rescued by restricting subjective assessment to those who are capable of making ‘autonomous’ and ‘informed’ assessments of their life, as Sumner (1996) does.

However, autonomy and being informed are unlikely to be clearly identifiable, and if they could be judged it would be upon a gradual scale rather than dichotomous criteria. There may be some cases that can easily be identified as failing to meet reasonable levels of autonomy and being informed, such as children, or people under positions of servitude. However, for

---

49 Drawn from a passage by Alexis de Tocqueville in which he notes that “man, in certain states, appears to be insensible to his extreme wretchedness” (de Tocqueville, 1935: Vol 1, Ch 18).
many, when autonomy is achieved is not clear. An individual may be inclinded to perceive their life as good because of social norms, or because they do not want to perceive themselves as a failure, or because it is psychologically conforting in the face of unchanging circumstances. Is an individual is such a situation autonomous? This shall be discussed in more detail below.

If impaired judgement includes ignorance about the feelings, emotions and life evaluation that the individual would make were they to have experienced a different life then every subjective judgement could potentially be impaired. For example, I may consider myself to be fully satisfied with my life in the absence of religious beliefs. However, having never been religious it is impossible to know if I would judge my life now as fully satisfied if I knew how satisfied I would be if I found religion.

Identifying those who make informed and autonomous judgements is practically problematic, and in most cases the criticism of potentially untrustworthy self-assessments remains. However, this criticism is likely to be tempered by the fact that most people are likely to feel they are in a better position to judge their life than an external person.

Daniel Haybron (2005) argues that life satisfaction judgments have a moral dimension reflecting virtuous ways of responding to our lives, and hence are subject to ethical norms. He argues that “our attitudes toward our lives can reflect various virtues and vices, such as gratitude, fortitude, ambition, pride, complacency, smugness, softness, low self-regard, etc.” (Haybron, 2005: 6). This suggests people may not be capable of making evaluative assessment without being influenced by their own value judgments on the appropriateness of different responses. If so, self-evaluation would be an inappropriate concept for interpersonal comparisons. People might tend to favor norms of gratitude as they get worse off, and norms of non-complacency as their situations improve (Haybron, 2005: 16). Consequently, “it could be reasonable for one’s life satisfaction attitudes systematically to have the opposite valence of one’s well-being” (Haybon, 2005: 7).

However, as seen in Chapter Three, life satisfaction usually does show the anticipated relationship with what we may think of as low welfare. The impact of moral beliefs about happiness on life satisfaction judgements is explored by Diener et al. (1991) who found a non-significant correlation between a morality of happiness scale and two SWB scales, suggesting that if individual morality influences life satisfaction judgements then this effect at least is minimal.
Interpersonal comparability requires that the meaning of different affects and evaluative questions is the same across individuals. However, cultural differences in the use of language may undermine cross-cultural comparisons of subjective measures. Goddard and Wierzbicka (1997) have argued that similar language can express different emotions within different cultures, suggesting both translation and conceptual differences. Some support for the cross-language comparisons of SWB has arisen through comparisons within multilingual countries. Veenhoven (1993) considers three tests to explore the impact of different language. Firstly, he finds different wording - ‘happiness in life’, ‘satisfaction with life’ and ‘the best/worst life’ - to give similar ranking between countries. Secondly, he finds less difference between language groups within countries than between shared language groups across countries. Thirdly, average happiness is not that similar between nations that use the same language, suggesting the language effect is not so strong that it dominates other sources of well-being. Individuals in the same language community have been found to share a common understanding of how to translate internal feelings into a numerical scale (van Praag, 1991). Where individuals are from the same linguistic and cultural background it may therefore be reasonable to assume a common meaning to common language.

Different cultures place different emphasis on the importance and appropriateness of striving for happiness. Individuals belonging to different cultural contexts may refer to similar levels of happiness yet “the components or attributes of the concept to which they refer are very different and the realization of each of them is equally different” (Galati et al., 2006: 608). Lu and Gilmour (2004) compare essays on ‘what is happiness’ from two cultural groups and find that Chinese students’ conceptualization of happiness incorporates more notions of self-control, harmony with the external world, less emphasis on positive hedonic experiences, and greater emphasis on social relationships than Caucasian American students (Lu and Gilmour, 2004). Similarly, satisfaction with happiness has been found to be a weaker predictor of life satisfaction in Hong Kong than in Australia (Lau et al., 2005).

Interpersonal comparability requires that different people use the scales, whether numerical or verbal, similarly. For the end points to be equivalent for life satisfaction questions, everyone must have equivalent conceptions of the worst and best levels of satisfaction, as discussed above this is problematic if end points depend upon past experiences. Kahneman (1999) argues that endpoints for states such as joy and anxiety are likely to be more consistent across individuals and less subject to manipulation due to changes in expectations and aspiration.

Furthermore, some people may be reluctant to express some types of affect to other people, or indeed to themselves. Similarly, there may be cultural influences which encourage/discourage
the expression and self-acceptance of some types of affect (Diener and Suh, 2000), giving rise to a situation where measurement error may be related to individual characteristics.

A further concern exists relating to whether people alter their true responses in order to give a socially appropriate response. For example, some groups may feel uneasy admitting to feelings of sadness, and may distort self-reports to present a favourable outward view. For example, Carstensen and Cone (1983) found a high correlation amongst the elderly between two frequently used measures of psychological well-being, and the Edwards Scale of Social Desirability (Edwards, 1957).

Whilst these tendencies may be consistent across one individual in different periods of time (hence not be problem for within person comparisons) they are problematic if people deliberately alter SWB responses to conform to socially acceptable responses following changes in circumstances. For example, the unemployed or recently widowed may report lower life satisfaction to meet social expectations. However, in the case of unemployment, past experiences of unemployment have also been identified as having a negative impact upon SWB when the individual is currently working (Lucas et al., 2004). This suggests that, unemployment has a negative impact on SWB.

Rather than response artefact, high correlations between social desirability scales and well-being scales appear to reflect content overlap between the scales (Diener et al., 2001). Diener notes that current social desirability scales “tap personality content which is actually related to SWB” (Diener, 1994: 114).

Subjective responses may also be influenced by the individual’s desire to appear happy to both others and themselves (Diener et al., 1991). However, a ‘happy image management’ scale was not found to correlate with SWB measures (Diener et al., 1991).

Some evidence to support the validity of interpersonal comparisons of affect scales can be taken from similarities between personal ratings and informant report ratings. For example, Costa and McCrae (1988) found that self-reports and spouse-reports were correlated 0.57 for positive affect and 0.49 for negative affect, over a six-year period. Similarly, Lepper (1998) reports self-other correlations of 0.43 to 0.45 for positive affect, 0.33 to 0.43 for negative affect, 0.47 to 0.53 for affect balance, 0.51 to 0.56 for SWLS and 0.57 to 0.64 for a composite variable of happiness, affect balance, and satisfaction with life. These high levels of cross-rater validity suggest individuals are able to recognize and predict the SWB of others hence

---

50 A scale measuring the tendency to give socially desirable responses in self-description.
51 Kimball and Willis (2006) argue that because life satisfaction questions invite the inclusion of past experiences, past unemployment may still affect responses due to how people think they should feel.
the language and the scales are being interpreted similarly across people. High cross-rater reliability has been found for interviewers, close friends and partners (Sandvik et al., 1993).

Support for interpersonal comparisons can also be taken from similarities between self-report SWB measures and physiological indicators of positive states. The level and location of brain activity has been found to be related to subject reports (Davidson 2000). For example, in right-handed people positive feelings are generally associated with more dominant brain activity in the left pre-frontal cortex, measured by electrodes on the scalp in Electroencephalography (EEG) or scanners in Magnetic Resonance Imaging (MRI). Other studies have also supported a link between specific brain activity and self-reported measures of affect and global life evaluations (Lane et al., 1997; Levesque et al., 2003; Urry et al., 2004).

Other physiological measures, such as salivary cortisol, have also shown links with self-reports of SWB. Ryff et al. (2004) found that people with higher SWB (specifically those with more meaning and purposeful engagement in their lives) have lower levels of cortisol and longer REM sleep. Positive affect has been found to be associated with reduced neuroendocrine, inflammatory and cardiovascular activity and inversely related to cortisol output (controlling for other factors such as age and gender) (Steptoe and Wardle, 2005; Steptoe et al., 2005). Pressman and Cohen (2005) reviewed studies relating positive affect to morbidity outcomes and found near unanimity across studies supporting the beneficial association of positive affect on morbidity, including prospective morbidity studies which found benefits of positive affect in conditions as diverse as stroke, re-hospitalisation for coronary problems, the common cold and accidents, suggesting a causal relationship. Lepper (1998) found that sleep disturbance correlated with negative affect, positive affect and affect balance, life evaluation and general levels of happiness, and that those with low negative affect and low affect balance had a higher prevalence of smoking. Associations have also been found between positive and negative emotions and startle eye-blink response and facial expressions (Ito and Cacioppo, 1999).

Additional support for interpersonal comparisons can be gained from the fact that determinants of SWB identified in cross-section (from variation between people, which picks up differences in levels) are similar to those identified using within-person analysis in panel data (which picks up the impact of changes). For example, married people have higher SWB and getting married increases SWB.

Despite concerns over threats to comparability, evidence suggests that comparisons of subjective measures are meaningful. The key to comparability rests on their being anchor positions within the scale that have identical meaning for all people, whether this be the best
and worst possible life, or some other anchor such as “happiness balanced by unhappiness” (Ng, 1980).

The assumption of rough interpersonal and inter-temporal comparability of most SWB measures is reasonable, subject to concerns over adaptation. One exception may be the GHQ scale. Interpreting the GHQ responses as interpersonally comparable is problematic since the questions compare the last few weeks to ‘usual’ yet what individuals are taking as their reference point is not clear. If a respondent repeatedly reports they are worse than usual, they may be saying that prior to a few weeks before the interview on each occasion they were better than they are now, or they may be using ‘worse than usual’ as a label for worse than their perception of what is usual in other people. Under certain circumstances a low score on the GHQ Likert scale may not represent lower well-being than a slightly higher score as intended, rather may just reflect greater stability. For example, the question ‘have you been able to concentrate on whatever you are doing?’ could be responded as ‘same as usual’ for someone who has on-going concentration problems, and ‘less than usual’ for someone with usually good concentration which has been below their norm in recent weeks. Consequently, assumptions of interpersonal comparability of this scale are problematic.

In the BHPS about a third of respondents who answer better than, or less than usual to the happiness question within the GHQ in wave 12 will also do so in wave 13 and of those responding much less than usual in wave 12 there is a 19 percent change they will also do so in wave 13. This suggests that some respondents are using the scale in a more absolute manner, hence presenting less of a threat to interpersonal comparability.

**Summary of interpersonal comparability**

In summary, no perspective is free from concerns relating to interpersonal comparability. Within preference satisfaction, ranking of the individual’s life is only interpersonally comparable if individuals use the same reference points. Where others rank alternative lives by imagining what it would be like to be the other person, if they also imagine the other person’s preferences where these preferences have resulted from adaptation to their circumstances, we may consider this to be the inappropriate perspective from which to make interpersonal comparisons.

Income as a proxy for preference satisfaction can only be interpersonally comparable where it is adequately adjusted to account for the diminishing returns to income. As also discussed in Chapter Two the comparison required is real income accounting for differences in price level and household composition which are not compensated for through other means.
Interpersonal comparisons of SWB measures are problematic if the subjective assessment is not trustworthy. As with preference satisfaction, there are concerns about the role of adaptation, where responses reflect deviation from the individual’s adopted norm rather than their absolute position. Concerns have been raised that SWB measures may be influenced by value judgements on differences in the importance of being content with one’s life circumstances, cultural differences in language and concepts, differences in how socially appropriate or desirable it may be to acknowledge and express certain levels of SWB. Despite these concerns there is growing evidence that interpersonal comparisons of SWB responses are meaningful. SWB measures correlate well with reports from others and physiological indicators.

5.3 Validity

As discussed in Chapter Four, the absence of a gold standard for well-being requires that validity assessments are made indirectly. Questioning certain properties of the measurement instruments should help determine whether we can confidently claim that the measure is a valid measure of individual well-being.

5.3.1 Content validity

To establish content validity we need to know whether the measure represents all facets of the concept, hence whether it covers all the important aspects of well-being adequately. The measure needs to be complete, in that knowledge of the measure is sufficient to provide knowledge of individual well-being.

Preference satisfaction

Preference satisfaction measured by income and consumption lack content validity. Were income to be a complete measure of well-being this would require a measure of income (real, permanent, possibly equivalent) to be the only item in an individual’s utility function. If other attributes are legitimately included, income alone cannot represent well-being. Leisure time, and non-market goods and services are typically included within an individual’s utility function at a theoretical level, but less likely to be included when it comes to actual measurement. More recent literature suggests that relationships should be treated as central to an individual’s utility. Lane for example, argues that in advanced economies friendship and a good family life are more important to well-being than income (Lane, 2000).

Actual preference satisfaction for some may lack face validity, hence the popularity of the saying ‘be careful what you wish for, lest it come true’. Income measures may have been seen to
lack face validity because people do not think money is the most important thing in their life (Defra, 2007), nor that it necessarily brings happiness. For example, Christianity and Buddhism present high income as a hindrance to their respective notions of well-being through the generation of attachment to materialism and possessions.

One means of indirectly establishing which factors are in a typical individual utility function is to assume that the level of preference satisfaction will follow SWB and consider the determinants of SWB. Studies have shown the income effect to be small (see Dolan et al., 2006a). Furthermore, some of the income effect is found to arise due to changes in relative income position (discussed further in Chapter Six) suggesting others income also has a place within an individual’s utility function. Studies have also shown that marriage, health and employment status and contact with friends and family are robust determinants of SWB (this is also supported in the findings in Chapter Three). These factors are independently important to SWB, and desires for them cannot be fully satisfied by increased income. Improved health status and achieving marriage or a stable relationship may be more likely following increased income, however, frequent contact with friends and family is less likely (as seen in Chapter Three). Consequently, the utility function appears to contain much more than just current income suggesting income alone cannot be a complete measure of well-being.

Although this looks like strong evidence that there is much more to life than income, because preference satisfaction and SWB measures offer two different accounts of what it is for a life to go well it is still possible to argue that people do in fact have higher well-being with higher income, even if they do not report higher SWB. Higher income could increase well-being even under circumstances where people feel no more satisfied with their life, or no more satisfied with their income. In other words, the additional attributes, marriage, contact with friends, employment etc. may matter independently for SWB but not for preference satisfaction.

Remaining within a preference satisfaction account of well-being we need to know whether factors in addition to own income influence preference satisfaction. The extent to which others income influences preferences has attracted considerable attention and has been widely debated (Scott, 1972). Relative income effects have been explored in three main ways. The first strand of empirical work uses experimental data on preferences towards lives with different income and relative income positions. The second strand uses laboratory games that involve choices over the allocation of money between the individual and a usually anonymous other. The third strand considers how concerns about relative status affect labour market outcomes.
The first strand includes a series of experiments where respondents choose which hypothetical scenarios would be in the best interests of an imagined future relative. For example, Johansson-Stenman et al. (2002) asked students to choose between a grandchild’s income of $2,500 per month when average income was $3,000, or a grandchild’s income of $2,300 when average income was $2,000. Respondent’s choices between future lives reveal both relative and absolute income concerns and the importance of positional concerns, particularly for visible consumption (Alpizar et al., 2005; Carlsson et al., 2003; Carlsson et al., 2005; Greene and Nelson, 2006; Solnick and Hemenway, 1998).

These studies suggest that when faced with hypothetical choices, people choose as though they believe that relative income will be important for how good life will be for the hypothetical person. However, Greene and Nelson (2006) critique these studies for being subject to framing effects because respondents are likely to know the study is about relative income, lack of effort because respondents have no incentive to provide the correct answer, and a dominant use of students who may have limited knowledge of what life is like working full time.

Furthermore, it is hard for respondents to envisage costs being constant between two societies. For example, if asked of UK undergraduates with plans to work in the city they would rightly deduce that limited housing supply in central London would result in them being priced out of the market and facing a long commute in an undesirable area if they do not give weight to relative income concerns. This would suggest a concern with relative income may arise from the impact relative incomes have upon real actual incomes through changes in the price of goods in fixed supply, rather than an independent concern with status or equity.

Although these choice experiments ask for an option which will be in a future individual’s best interests, they may be influenced by what the respondent thinks would make a better society or by moral satisfaction from choosing a more virtuous option (Kahneman and Knetsch, 1992).

Experiments on sharing money suggest that preferences are influenced by the amount of money others have, some suggesting income of others enters positively into an individual’s utility function (Charness and Grosskopf, 2001; Fehr and Schmidt, 2006; Forsythe et al., 1994; Guth et al., 1982, Hoffman et al., 1996; Kahneman et al., 1986), others suggesting it enters negatively (Zizzo and Oswald, 2001). However, the extent to which preferences towards fairness in the income distribution, procedural fairness or relative income position influence results is not clear.
The consideration of relative income effects on labour market outcomes has found that relative concerns impact upon the decision of females to join the labour market (Neumark and Postlewaite, 1998; Woittiez and Kapteyn, 1998), decisions to quit (Brown et al., forthcoming; Kwon and Meyersson Milgrom, 2006), and the distribution of wages (Frank, 1984).

These three research strands contribute towards a general picture that others’ income is important to individuals and influences their preferences. If an individual’s utility function contains attributes in addition to own income, then well-established methodologies for inferring a monetary value to non-market goods can be used to value these non-income attributes, adjusting income to allow for the value of all relevant non-market goods consumed by the individual. However, practical difficulties have prevented a fully adjusted individual consumption measure to be estimated.

Measures which focus on HRQoL will lack content validity if they are taken as a complete measure of well-being if other things matter to well-being. Only physical and mental health are included in the descriptive systems used in QALY measures. However, it was argued in Chapter Three that since the valuations of health states are derived from trading off changes in health against sacrifice in number of years of life (TTO) or the probability of sudden death (SG), this implies that everything of value is incorporated within the health outcome measure.

Subjective

Subjective measures of well-being have been criticised for excluding the true state of the world in addition to our subjective assessment and response to the perceived state of the world. For example, Harsanyi (1982) notes that, “It is at least arguable that in many cases we are more interested in achieving some objective state of affairs than we are interested in our own subjective feelings of pleasure and pain that may result from achieving it” (Harsanyi, 1982: 54).

The objection that hedonism does not capture all that is of value is captured famously by Robert Nozick’s experience machine. Nozick claims that many would be reluctant to “plug in” to an experience machine “that would give you any experience you desired” (Nozick, 1974: 42-3) implying a concern with actually doing things and having a real experience in addition to wanting the feelings associated with doing. The thought experiment suggests that people value truth and realness of experience in addition to the felt and subjective experience and care about states of the world, beyond their feelings. Although there may be other explanations for our reluctance to plug in (lack of trust in the reliability of the machine, for
example) subjective measures struggle to overcome the criticism that most people’s concerns “are not limited to the quality of their own mental states” (Scanlon, 1991: 23).

Subjective: Hedonist

Hedonist measures have also been criticised for excluding subjective attributes beyond positive and negative affect. It is known that aspects of evaluation and assessment vary from positive and negative affect (Diener, 1994), which casts doubt on the completeness and content validity of feeling-only measures of well-being. Measures of affect may not incorporate all affects which matter to well-being, for example they usually exclude engagement or ‘flow’ (Csikszentmihalyi and Hunter, 2003). Hedonist measures of positive and negative affect may miss some of the benefits arising from the experience of negative emotion, for example, individuals may perceive that they benefit in some other way from feeling negative affective experience. Melancholy or appropriate grief may be interpreted in a positive light.

Hedonism has also been criticised as missing something important to human existence by weighting all pleasures equally, hence pleasures of the ‘swine’ are given equal weight to cultural and aesthetic pleasures. Bentham famously argued that pushpin and the arts and sciences should be valued in proportion to the pleasure they yield (Bentham, 1843). Mill on the other hand pointed to the existence of higher level pleasures (literature, art, contemplation, creativity, activities which require initial effort), which he claimed should be given greater weight than simple pleasures. If another quality in addition to the quantity of pleasure experienced matters to well-being then simple aggregation of pleasures will lack content validity.

However, an intuition favouring higher pleasures may arise because these ‘higher’ pleasures have longer lasting effect than those commonly perceived to be lower pleasures (such as sex, eating, and scratching). The former appeal to many different aspects of the self, their complexity resulting in future benefits arising as the individual is able to appreciate the world in a different way following the experience. Hence any intuitive sense of some pleasures being ‘better’ than others could be a question of maximising long-term aggregate pleasure.

Crisp raises a related problem that if lifetime well-being is the aggregate of moments, then there should come a point where a very long life of many moments of low well-being (he gives the example of a barely sentient oyster) will become preferable to a shorter life of higher moments of well-being (the life of Haydn in his example) (Crisp, 2006). This seems problematic if our intuition would favour the shorter, intense life.
Limited life expectancy may allow us to sidestep the problem but does not address the underlying issue, or cause of our intuition that a short, meaningful life may be preferable to a long, pleasant but non-meaningful life. Our intuition may lie in desires for attributes in life beyond our own well-being, such as morality and virtue. However, being barely sentient does not fit well with the notion of life going well for the individual. This may be because particular value is given to intense experiences, undermining the ability to aggregate experience over time.

Theoretical criticism levied against hedonism highlights that important attributes such as meaning, self-assessment, self-identity, values, and beliefs may not be being adequately captured within an aggregation of positive and negative affect. However, we have little knowledge of the extent to which states of the world, or experiences and evaluations such as meaning and self-assessment, which we may intuitively feel to be important beyond our affective experience, are not in fact captured within affective experience. We may intuitively feel certain attributes are non-instrumental (such as meaning) even if they are in fact instrumental to maximising felt pleasure.

Although hedonism may be enjoying some resurgence, the criticism that it cannot capture everything that we care about, for our own sake, seems unlikely to go away. Empirically it is not possible to assess how much, if anything, is missing from hedonist measures of well-being, hence our intuitions can not be adequately tested. However, whilst such intuitions are voiced measures based on hedonism will be the target of objections that will be difficult to defend.

Subjective: Evaluative

From the perspective of an evaluative account of well-being, an individual’s assessment of their life has the potential to incorporate everything of value for that life that they are aware of. This can include aspects of life that extend beyond feelings, such as evaluation of self-identity and beliefs. However, it will remain subjective, therefore, still face criticism that the truth of the matter is something we care about beyond our own experience of the world.

If an individual claims to be extremely happy and satisfied with life, but is not educated, has no material resources, has low income, no close relationships and is in very poor health, then we may find it difficult to accept their own assessment of their life as a good proxy for their well-being. This may stem from three reasons, firstly, a belief that certain objective states are sources positive affect and positive life appraisals therefore some individuals are misguided in their assessment. Secondly, we may care about objective states because they are intrinsically
good regardless of whether they are evaluated as such by the individual. Thirdly, our concern may be less about well-being than justice and responsibility, for example, we may think that those with a high income should be considered to be more responsible for their well-being than those with a low income (Cohen, 1993). Elster (1983) explains this as not wishing to punish those who adjust to circumstances with excessively low aspiration, nor reward those who have unrealistically high expectations.

If concern arises from the last reason, this does not necessarily threaten the validity of the measure of well-being, but creates a normative argument for weighting the contribution of different individuals’ well-being within a social welfare function. The second reason suggests adherence to a substantive good account of well-being and that individual assessments will not have content validity because important aspects will be overlooked. The first reason, that assessments are misguided, suggests problems with making interpersonal comparisons discussed above.

If the measure of well-being uses an aggregate of domain satisfactions then content validity may be questioned if domains are excluded which are perceived by others to be important, or shown to be related to overall assessments through data analysis.

Face validity requires the measure to be relevant for the population to which the measure is to be applied. This is likely to be particularly important for domain satisfactions. For example, if a domain-weighted satisfaction measure incorporated job satisfaction this lacks face validity for children and the retired.

**Summary of content validity**

Different strands of empirical work have shown fairly conclusively that individuals’ utility functions do not contain absolute income alone, suggesting income lacks content validity as a measure of well-being. Mental state measures will all suffer from a concern that well-being contains the truth of the matter rather than perceptions of circumstances. However, broader evaluative measures are less subject to criticisms that important content is excluded compared to affect only based measures.

**5.3.2 Convergent validity**

A measure with convergent validity should correlate highly with theoretically similar measures. Whilst there are theoretical differences between the measures, they all purport to measure well-being, hence their outcomes should be broadly similar. Measures of well-being should correlate with physical signs of well-being such as observer reports. We should also
expect measures to respond to situations that can be taken, fairly unambiguously, to detrimental to well-being, such as non-voluntary social isolation, unemployment, and poor health.

Preference satisfaction

As discussed in Chapter One preference satisfaction measured by income converges with some aspects of well-being such as education and health. However, in other areas, such as happiness and depression amongst the young, there is little relationship. Income may converge with other indicators of well-being at some levels of income better than others. For example, income correlates more highly with other well-being outcome measures at low levels of income than high levels (Diener and Seligman, 2004; Ecob and Davey Smith, 1999).

Subjective

The evidence supporting comparisons of subjective measures across individuals also offers some support for the convergent validity of these measures.

Evaluative measures, such as life satisfaction, converge with other methods of well-being measurement, such as third party reports of their life satisfaction, frequency of smiling, daily mood ratings, and the number of positive and negative events recalled (Pavot et al., 1991; Seidlitz and Diener, 1993, Sandvick et al., 1993; Diner and Suh, 1997; Diener, 1994). Various physiological and neurological phenomena show strong correlation between emotions and facial expressions, smiling, frowning, and SWB questions (Sandvik et al. 1993).

Health is a strong determinant of life satisfaction (as shown in Chapter Three). However, in most studies health measures are subjective, hence subject to endogeneity problems. Stronger evidence of convergent validity can be taken from studies that use objective outcomes of health, such as suicide, risky behaviour and blood pressure.

Bray and Gunnell (2006) demonstrate that suicide is significantly negatively correlated with life satisfaction and happiness in the EVS 1999/2000, although the findings from cross-section data on Western nations are not very robust (Di Tella et al., 2003). Veenhoven (1993) found a strong correlation between country level measures of anxiety behaviour (e.g. consumption of stress related stimulants) and overall happiness.

Blanchflower and Oswald (2007), using 2001 Eurobarometer data, relate differences in life satisfaction across countries to differences in self-reported high blood pressure problems and find a strong inverse relationship. The pattern of the country dummies in a regression predicting blood pressure problems is very similar to the (inverse of the) pattern of the
dummy variables for countries in a life satisfaction regression. This finding is cross checked against life satisfaction and happiness data from the ESS (2002 & 2004) and using six questions from the GHQ12 in the Eurobarometer data. High blood pressure enters with a significantly negative coefficient in an ordered regression predicting life satisfaction in the Eurobarometer data and using data from the 1999/2000 British National Child Development Study.

**Summary of convergent validity**

The convergent validity of affect measures and SWB is well supported with a wide range of studies. In terms of income, less research effort has been dedicated to showing that income converges with attributes that we associated with well-being, in part due to a lack of questioning over those relationships. However, the lack of convergence with measures of SWB over time brings convergent validity of income measures into doubt.

### 5.3.3 Predictive validity

Outcome measures will show predictive validity if they are able to accurately predict behaviour or future states that are expected to relate to the outcome.

**Preference satisfaction**

As discussed in Chapter One income predicts a range of positive outcomes such as health, life expectancy, and education performance.

**Subjective**

**Subjective: Hedonist**

The links between positive affect and improved health states can be seen as both evidence for interpersonal comparability, convergent validity and predictive validity. There is some evidence of a link between current positive and negative affect and future health states, including risk of hypertension (Jonas and Lando, 2000; Steptoe and Wardle, 2005), future stroke (Ostir et al., 2001), and mortality for community-dwelling older people (Pressman and Cohen, 2005).

**Subjective: Evaluation**

Evaluative measures have also been shown to predict behaviour, such as reduced suicide attempts (Moum, 1996, cited in Diener, 2000) and both intentional and unintentional fatal
injury, independent of baseline health (Koivumaa-Honkanen et al., 2002). SWB measures also correlate highly with morbidity and mortality outcomes. A high correlation has been found between low SWB and coronary heart disease (Sales and House, 1971), and length of life (Danner et al., 2001; Deeg and Zooneveld, 1989; Huppert and Whittington, 2003; Palmore, 1969), particularly for poorer countries (Veenhoven, 1993). Those with high life satisfaction are less likely to catch a cold when exposed to the cold virus and recover more quickly (Cohen et al., 2003). Life satisfaction has also been shown to predict future marital break up (Gardner and Oswald, 2006).

Confidence in other subjective measures has been gained through their predictive ability. For example, subjective measures of health have been shown to predict both suicide (Helliwell, 2003) and mortality (Idler and Benyamini, 1997). Subjective assessment of job satisfaction predicts future quitting even controlling for wages, hours worked and other variables (Clark et al., 1998, Clark, 2001; Kristensen and Westergaard-Nielsen, 2004). Clark (2003a) finds that mental stress scores on entering unemployment predict unemployment duration and those who suffered the largest drop in SWB on entering unemployment were quickest to leave.

It might be expected that individuals would choose to discontinue activities associated with low well-being (Frijters, 2000; Shiv and Huber, 2000). However, change requires more than an incentive to change, therefore the absence of a relationship between low well-being and changes in circumstances does not necessarily undermine the validity of the well-being measure.

**Summary of predictive validity**

Predictive validity is a useful way of showing that the measure is picking up real differences between people, and that it is measuring something that has genuine meaning. However, the direction of causality is not always clear, nor how the level of well-being should theoretically impact upon behaviour. For example, the level of unhappiness amongst unemployed may be hypothesised to influence their behaviour, however, happier unemployed may be more likely to remain unemployed because the state is relatively less bad for them, alternatively happier unemployed may have more motivation to find new employment. Hence while we would expect levels of well-being to predict future behaviour and future states it is not obvious exactly what we would expect high current levels of well-being to predict.
5.4 Empirically useful

5.4.1 Reliable with low and unbiased measurement error

It is assumed that well-being is measured with some error, such that reported well-being $WB_i$ is true well-being ($WB_i^*$) plus measurement error ($\varepsilon_i$)

$$WB_i = WB_i^* + \varepsilon_i$$

The usefulness of any measure for public policy will depend partly on the size of the measurement error, but more crucially on whether the measurement error contains systematic biases.

One method for testing the size of measurement error is to consider the reliability of instruments in sense of repeatability or consistency over a period of time during which it is assumed true well-being will not have changed. Instrument reliability can be tested by test-retest correlation. This should offer information on the size of random measurement error, however, a measure may be highly reliable and at each time period be subject to the same systematic measurement error.

Another method is to consider whether unrelated factors influence the measures.

Preference satisfaction

Income as a measure of preference satisfaction is subject to concerns of the reliability of self-report, particularly where people perceive an incentive to misrepresent their income. Reliability issues are particularly acute within the self-employed and informal sector (Moore et al., 2000). However, in general income measures show high reliability, typically over 90% (Krueger and Schkade, 2007). There is some evidence that the reliability of stated WTP responses are reasonable (McConnell et al., 1997; Carson et al., 2003) with reliability increasing in line with WTP values (Smith, 2007), possibility due to the additional thought involved.

Chapter One raised a variety of ways in which our choices have been found to conflict with utility maximising behaviour, which would imply considerable measurement error. There are good theoretical reasons supported by an increasing body of empirical evidence to suggest why choices may not reflect underlying preferences, and may result in less of an individual’s preferences being satisfied. The fault rests not with preference satisfaction as a theory of well-being but with the use of choice (and WTP) as a reflection of preferences. However, whilst
research has clearly shown that choices can be misguided, the important question is to what extent are they likely to be misguided for the average person in ways that matter for the use of revealed preference as a measure of well-being. More specifically, the extent to which individuals’ choices in relation to income generation and consumption diverge from utility maximising behaviour.

Turning to health-related utility measures, the reliability of descriptive systems for multi-attribute health-related utility scales such as the SF-6d and EQ-5D has been found to be high. For example, test-retest reliability over a 2 week period for patients with Multiple Sclerosis was found to be 0.83 for the SF-6d and 0.81 for the EQ-5D (Fisk et al., 2005). Test-retest reliability has also been shown to be high for the SF-6d in patients with systemic sclerosis: 0.82 over an average of 5 weeks (Khanna et al., 2007). However, less is known about the reliability of the preference weights that support these measures.

Subjective

Studies reporting the reliability of subjective measures show a wide range of test-retest correlations, but in general show considerable stability (see Table 20 in Appendix 5.A). This level of stability would be unlikely if SWB measures were dominated by measurement error. The overview of the research presented in Table 20 and the meta analysis conducted by Schimmack and Oishi (2005) shows that multiple item scales produce stronger retest correlations than single-item measures, especially over short time intervals. As would be expected, the test-retest correlation declines overtime.

SWB responses (evaluative and hedonic measures) may include error from the reluctance of individuals to disclose information about their life, social desirability, self-presentation based on moral beliefs about happiness, and happy image management. In the discussion earlier these sources of measurement error were shown to be within acceptable limits.

Hedonic measures may also be subject to measurement error if individuals are not themselves aware of their current affective state. For example, Haybron argues that feelings like anxiety or tension may not be noticed even when they are occurring (Haybron, 2007).

Hedonic measures and reports of affect may also be subject to recall error. Experienced affect may vary according to when the assessment occurs. If an assessment of affect is conducted at the moment of experience, this may result in a different subjective assessment than if it is conducted some time after the experience (see Kahneman, 1999). If true affect is that which is experienced at the time, then problems with memory recall introduce a source of measurement error.
Chapter 5  How closely do well-being measures meet the ideal criteria?

Evaluative measures may be subject to measurement errors if individuals do not have a clear knowledge of their evaluation of their life. However, this should not generate any systematic measurement error unless responses are unduly influenced by what the respondent’s attention is drawn to at the time of assessment. Answering a life satisfaction question requires a difficult mental task and it has been proposed that respondents are unlikely to retrieve all the information that may be relevant to an overall assessment of their lives. Instead, they are likely to truncate the search process when they think they have enough information to form a judgement (Schwarz et al., 1991). Schwarz and Strack claim that SWB responses are “better conceptualised as the result of a judgement process that is highly context-dependent” (Schwarz and Strack, 1999: 62).

If current affect unduly influences life satisfaction assessment, this generates measurement error correlated with current mood (Schwarz and Clore, 1983). There is some evidence, for example, that respondents reported being more satisfied with their lives on sunny days than on rainy days – but only when their attention was not drawn to the weather (Schwarz and Clore, 2003). In addition, current affect may activate memories associated with that mood (Bower, 1981); for example, when people are put in a good mood they have a better memory for happy events or positive words (Argyle, 1987: 138). However, in natural settings mood, and consequent error, is likely to be random, unless mood itself was related to an external stimulus (such as success in a football match, or unusual weather) which varied systematically across the sample. The effect of mood may impact more strongly on overall life satisfaction than domain satisfactions since the criteria for evaluating domains are more clearly identified (Schwarz and Strack, 1999). Diener et al. (1991) examined the effects of current mood on SWB and found that the average correlation of three time periods for the Fordyce scale of SWB was 0.67, whereas average correlation of mood was only 0.33 suggesting SWB is more stable than mood. The highest correlation was not always period 1 SWB and period 1 mood, suggesting that high correlation arises because of a long-term relationship between the two. Regression analysis with the Fordyce score as the dependent variable and a SWB component score and current mood as independent variables, found that current mood can sometimes have a significant relationship on single assessment of SWB, beyond the relationship with long-term SWB, although the influence is slight.

A commonly examined context effect is the influence of the preceding question. As described by Schwarz and Strack responses are “subject to pronounced question order-effects because

---

52 Happiness on an 11-point scale anchored at extremely unhappy (utterly depressed) and completely happy (feeling ecstatic, joyous, fantastic!).
the content of preceding questions influences the temporary accessibility of relevant information” (Schwarz and Strack, 1999: 79). However, meta-analysis of eight articles reporting 16 studies that manipulated item-order to test the influence of temporarily accessible information generally showed heterogeneous findings (Schimmack and Oishi, 2005: 37). They find specific-global correlation to be significantly higher than global-specific, but the average effect size is weak to moderate. Studies by Strack et al. (1988) using ‘happiness with dating’ and ‘happiness with life as a whole’ found the strongest order effect, however, this finding was not repeated in a replication of the study (Schimmack and Oishi, 2005). These question order effects may be mitigated by ensuring global questions are asked at the beginning of surveys.

Information about how well life is going may be made accessible by other current events, such as what is in the news, what has happened that day, or where the survey takes place. Responses have been found to be related to the outcome of a football game (Schwarz et al., 1987), finding a dime, the presence of a handicapped person in the room (Schwarz and Strack, 1999), and the pleasantness of the room in which testing occurs (Schwarz et al., 1987).

Context effects limit the comparability of results obtained in different studies, and cast doubt on the clarity of meaning and our understanding of the life satisfaction measure. This is seen by some as undermining the use of SWB measures as a tool in public policy (e.g. Schwarz and Strack, 1999: 80).

An important question is whether incidental factors (such as weather, current mood, and question ordering) are likely to create sufficient measurement error to undermine the calculation of a non-bias value for SWB. Of course, what constitutes incidental is a tricky question. One option may be to encourage respondents to reflect on whether their response has been unduly influenced by a particular factor. However, many of these problems may be potentially over rated as many well-cited studies have not been replicated.

*Summary of reliability and measurement error*

Income measures fare best in terms of reliability, but reliability of subjective measures is not of a level that would undermine their use in well-designed surveys. All measures face many theoretical reasons for measurement error, and potential sources of systematic bias. However, those facing subjective measures appear minor in comparison to those facing income.

53 The order of questions about ‘happiness with dating’ and ‘happiness with life as a whole’ changes the correlation to SWB from 0.16 when the general happiness question preceded the dating question to 0.55 when the order was reversed (Strack et al., 1988).
Although income measures are a reasonably accurate measure of income, they are a poor measure of preference satisfaction.

### 5.4.2 Cardinality

To enable aggregation and estimation of average well-being across different groups it is preferable to have a measure which gives a single value and can be confidently treated as cardinal. At the very least we require the measure to be ordinal such that higher levels will unambiguously represent higher well-being.

**Preference satisfaction**

Preference satisfaction does not lend itself well to a single value scale, nor to cardinalisation, nor to natural anchor points. Knowing that having more of one’s preferences met enhances well-being does not say how much well-being has been enhanced. However, if the intensity of all held preferences can be compared, then this may offer a potential means for weighting across different preferences by intensity. If WTP is an indication of intensity of preference, this leads to the potential for income to represent intensity-weighted preference satisfaction. This would provide a single value of the estimate of the anticipated change in preference satisfaction. However, the above two premises on which this conclusion rests are hard to justify, as shown in the following discussion.

The absence of a one single measure of income or consumption undermines the estimation of a single value to reflect preference satisfaction. Preference satisfaction proxied by consumption of market and non-market goods will lead to different levels of well-being arising because of potential disagreement on which non-market goods should be included and different methods of valuing non-market goods. For example, should living in a safe neighbourhood be incorporated and if so, how is it to be valued (this has been done in the past by both hedonic pricing methods (Gibbons, 2004) and stated preference methods (Atkinson et al., 2005)). Should consumption of ‘regretables’ be included, such as cost of commuting to work, or is this a choice variable?

As discussed in Chapter Two, if income is used to represent preference satisfaction it will vary depending upon whether individual or household income is used, the method by which income is allocated to household members, the method by which nominal incomes are translated into real incomes, whether annual or current income is used, the extent to which local and national taxation and subsidies are incorporated, and whether income is net of saving and dis-saving (and the method used for assessing changes in wealth). There is no
single income measure, and good reasons for supposing these different income measures may show considerable variability (as seen in Chapter Three).

If a single scale of intensity of preference satisfaction could be derived, based on WTP, would this be cardinal? When measured in monetary terms, ‘utility’ gives the appearance of being cardinal. However, as discussed in Chapter One it is widely accepted that there are diminishing marginal returns to income, hence income can no longer be treated as cardinal. However, if we can adjust income for the diminishing marginal utility of income, such that one unit increase on the adjusted income scale represents the same intensity-weighted desire satisfaction for all levels of income, then the adjusted measure of income may be cardinal. As discussed in Chapter One, there is considerable disagreement on the elasticity of marginal utility of income.

Despite the potential to adjust income to account for the satisfaction of non-market preferences, in reality an income measure of well-being is unlikely to incorporate the fulfilment of non-market desires. Income will only measure satisfaction of market-based desires. To the extent that they are not perfectly correlated with satisfaction of non-market based desires, income as a cardinal measure of desire satisfaction will be undermined. Moreover, if non-market bads (such as the fear of burglary, stress at work, reduced contact with friends and family) are positively correlated with market consumption (and income), the relationship between income and preference satisfaction may not even be clearly ordinal.

In summary, if theoretical agreement was reached on exactly which measure of income reflects preference satisfaction, and we were confident that non-market desire satisfaction did not pose a substantial problem, and we knew the elasticity of the marginal return to income, it would be possible to assume that adjusted income could offer a single value which could be treated as roughly cardinal.

Preference based QALY measures are cardinal by assumption, such that a change from 0.4 to 0.5 is equivalent to a change from 0.8 to 0.9. The cardinality of the bounded (0-1) QALY weights is derived either from changes in risk (SG), time neutrality (TTO), or an assumption that individuals treat rating scales as cardinal (VAS).

Whilst a QALY weight is always given the same meaning, there are different methods for generating those weights, suggesting the same health outcome could be assigned a range of possible QALY weights. Variation arises because there are different health state descriptive systems which may be used (e.g. the Sf-36, EQ-5D), the valuation of each health state may vary depending on whose preferences are used (e.g. patients or the public) and the methods of
valuation used may vary (e.g. TTO, SG, VAS). Hence although cardinal by assumption there is variability in the outcome measure according to different techniques and assumptions used in generating the measure.

*Subjective: Hedonist*

Hedonism assumes a single metric can value well-being, that of ‘pleasurableness’, or ‘enjoyment’ (Crisp, 2006) consequently is well suited to a single outcome measure. Ideally we would like each experience to match one level of the outcome measure. However, measures of positive and negative may vary if different types of affect are collected, if intensity is measured in different ways, and if affects are aggregated in different ways.

Hedonist measures require combining different types of positive and negative affect and intensities of affect into an aggregate level of ‘pleasurableness’ experienced in a moment of time. A cardinal scale for ‘pleasurableness’ of the moment can be generated if the means to weight different affects and account for differences in intensity are established. If time is treated neutrally then this creates the means by which cardinality of a measure of well-being over a period of time can be established. However, any ambiguity on whether the presence of particular affect contributes or detracts from the ‘pleasureableness’ of a moment (such as experiences of melancholy) may undermine a clear order of moments.

One problem arises if positive and negative affect are not part of the same spectrum. As discussed in Chapter Two this remains an area of controversy. If positive and negative affect are found to be different dimensions, the method by which these are combined needs to be clearly established. Ad hoc aggregation according to the number of positive and negative affect questions will not be theoretically consistent. Where internal consistency checks suggest more than one dimension, the means by which these different dimensions are aggregated should draw upon additional theory. Adhering to preference hedonism would enable weights that reflect overall perceived importance of different affects to an individual’s meta ‘pleasurableness’ scale.

*Subjective: Evaluative accounts*

Evaluative accounts measured by single item questions offer a unique measure of well-being. However, the exact wording of the evaluative question, from life satisfaction to overall happiness or contentment, introduces variability. Multi-item questions, whether focused on overall evaluation (e.g. SWLS) or as an aggregate of domain satisfactions (e.g. PWI) introduce potential variability through the number of possible options for combining items.
SWB measures involving aggregation of a number of questions and those that draw from a combination of accounts have no natural unique scale. Measures that proxy for SWB may give quite different levels of well-being, particularly if there is no a priori theoretical judgement as to the contribution of one aspect (such as positive affect) compared to another aspect (such as positive evaluation). However, the measures may be aggregated into single scales that may artificially be treated as cardinal.

When respondents are faced with single item evaluative measures, say a life satisfaction on a scale of 0 to 10, they may use what would seem a reasonable assumption that the scale is linear, with equal distance between each level. However, strictly, we only know that the scale is at least ordinal and it is possible that reported life satisfaction ($LS_i$) is a non-linear function of true life satisfaction ($LS_i^*$), such that $f$ is a non-linear function.

$$LS_i = f(\text{LS}_i^*)$$

When people are faced with a bounded scale, responses may be compressed at either end. This has been found to be the case in the measurement of pain (Hartmannsgruber et al., 1999).

Suppose that an individual answers 6 on a 1-7 life satisfaction scale. There are many interpretations of what this answer could mean. ‘6’ could mean the individual’s actual life satisfaction is between 5.5 and 6.5, 6 and less than 7, 5 and 6, or 6 only. Table 19 outlines some of the possible interpretations.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.5 to 1.5</td>
<td>1.5 to 2.5</td>
<td>2.5 to 3.5</td>
<td>3.5 to 4.5</td>
<td>4.5 to 5.5</td>
<td>5.5 to 6.5</td>
<td>6.5 to 7.5</td>
</tr>
<tr>
<td>(2)</td>
<td>1 to 1.5</td>
<td>1.5 to 2.5</td>
<td>2.5 to 3.5</td>
<td>3.5 to 4.5</td>
<td>4.5 to 5.5</td>
<td>5.5 to 6.5</td>
<td>6.5 to 7</td>
</tr>
<tr>
<td>(3)</td>
<td>1 to 2</td>
<td>2 to 3</td>
<td>3 to 4</td>
<td>4 to 5</td>
<td>5 to 6</td>
<td>6 to &lt;7</td>
<td>7 only</td>
</tr>
<tr>
<td>(4)</td>
<td>1 only</td>
<td>&gt;1 to 2</td>
<td>2 to 3</td>
<td>3 to 4</td>
<td>4 to 5</td>
<td>5 to 6</td>
<td>6 to 7</td>
</tr>
</tbody>
</table>

Note: Equality signs are not included in this table since there is no reason for thinking them to be in either direction, this strictly adds an additional possible row in each case.

Because the scale is bounded between 1 and 7 respondents may restrict their judgement to within the scale given, which does not go below 1 or above 7 (as in row 2), resulting in the top and bottom categories being relatively underused. This is more problematic for the top end since the life satisfaction distribution is strongly negatively skewed.

Alternatively, respondents may perceive either end of the scale as representing only the extreme values, and only answer 7/7 when they perceive their life such that no improvement
was possible (as in row 3). Consequently, it is not clear whether 7/7, for example, means 100% of the scale (row 3) or somewhere in the top seventh of the scale (>0.857) (row 4).

Van Praag (1991) offers some evidence supporting the possibility of cardinal interpretation, finding that respondents use five verbal labels such that they may be interpreted using an ‘equal quantile assumption’. The cardinal assumption is likely to be more relevant when numerical scales are used rather than verbal labels (Cummins and Gullone, 2000).

The cardinality of numerical scales of life satisfaction and happiness questions is explored by Layard et al. (2007) who conduct three tests to investigate the linearity of the life satisfaction/happiness reporting function.

The first test is a spline test conducted on six datasets in which the predicted values from the happiness regression are regressed on reported values separately for those with happiness levels above and below the median. They find that at low levels of happiness the predicted values underestimate reported happiness, yet at high levels of happiness the predicted values overestimate reported happiness, consistent with a non-linear transformation. However, this test relies on the assumption that the correct model is being used. As seen in Chapter Three, the relationship between education and SWB differs between the top and bottom end of the SWB distribution. This test would not distinguish between the scale being treated as non-linear, and the relationships between attributes and SWB varying at different points of the SWB scale.

Secondly, they consider the individual average variance of the residuals from panel data to see whether residuals have a lower variance at high average happiness values. Using the GSOEP they find that average root mean square predicted error is higher for low individual average reported happiness. However, as the authors note this may point to fewer mood swings among people at the upper end, or heteroskedasticity in the error term.

Thirdly, they look at ordered probit models and find significantly improved fit as compared with OLS, suggesting a non-linear transformation between utility and reported happiness. They also consider ordered logit cut points finding them slightly convex with respect to reported happiness (implying a concave transformation from true utility to its reported values).

Despite some evidence of a non-linear transformation, Layard et al. (2007) find that adjusting to account for non-linearities makes little difference to the estimate of the elasticity of marginal utility of income. A noted previously, others have also found that assuming cardinality or ordinality of the responses to life satisfaction questions is relatively unimportant.
for the results on the determinants of well-being (Ferrer-i-Carbonell and Frijters, 2004, Frey and Stutzer, 2000).

Where single response questions are not strictly cardinal, yet the differences between the intervals are not too variable, it appears that the simplification to cardinality is reasonable, and the advantages, particularly that of enabling the use of means, are likely to outweigh remaining concerns over lack of perfect cardinality.

Summary of cardinality

If there are diminishing returns to income then income must be adjusted before it can be treated as a cardinal measure. Even taking a fairly conservative estimate for the elasticity of the marginal returns to income of one, as recommended by the Treasury, this implies a considerable adjustment to income. This also implies GDP per capita would bear little resemblance to average well-being. There seems no great impediment to treating numerical subjective measures as cardinal but awareness of this simplification should remain apparent.

5.4.3 Appropriate sensitivity

A measure of well-being will only be useful for public policy if it is sensitive to changes in factors over which society has some control. Sensitivity can be considered partly through consideration of the range of variation within the measurement scale. An increase in the number of response options would be expected to enhance sensitivity (see Russell and Bobko, 1992), particularly if respondents use the full range of the scale. An indication of sensitivity to changing circumstances can be gained by considering the stability of the indicator over time, on the assumption that external circumstances are likely to change with time. More directly, sensitivity can be shown by evidence of change following known changes in circumstances that are thought to impact upon well-being.

Preference satisfaction

Income and consumption are potentially the most sensitive measures since it is possible to show very small changes, and they are both unbounded. In practice, because of a general reluctance to report income, many studies gather income data in ranges, hence reducing sensitivity. Nevertheless, this does not necessarily imply sensitivity to true changes in well-being.
The degree of sensitivity of a QALY measure will depend on the descriptive system used, however, since this descriptive system aims to cover all HRQoL it is unlikely to be sensitive to small changes across one dimension of health or to changes in non-health aspects of life.

*Subjective: Hedonist*

Hedonist measures, such as the ESM and DRM, have the potential to be highly sensitive to even short-term changes in well-being. Recent versions of the DRM give reports on roughly 12-14 episodes a day where each episode is evaluated in terms of 8 dimensions on a six point scale hence sensitivity is potentially extremely high. If affect has strong state like qualities (e.g. cheerfulness may be a property of the individual showing little variation over time), then these measures may be relatively insensitive to changes in circumstances. However, measures of net affect show considerable variation across activities, suggesting circumstances do matter (Kahneman et al., 2004b).

*Subjective: Evaluative*

The degree of sensitivity of life satisfaction questions will depend upon the scale used. Cummins and Gullone (2000) note that a Likert scale of 5 to 7 response options does not exploit the discriminate capacity of most people, particularly as responses are strongly influenced by trait characteristics and that the majority of respondents will only use the top end of the scale. Consequently, they argue that an 11-point (0-10) scale is preferable for attaining maximum sensitivity with no loss of reliability. Multiple item questions, such as the SWLS, have greater potential variability.

At present, analysis of large datasets has only explained a small proportion of the variation in life satisfaction in terms of an individual’s circumstances, indicating either fairly rapid adaptation to new circumstances or that currently identified external circumstances play only a small part in determining life satisfaction. For example, Argyle (1999) estimates that external circumstances account for about 15% of the variance in SWB.\(^54\)

Evaluative measures have generally been found to be fairly stable over time, whist a positive sign in terms of reliability, does suggest limited sensitivity. Despite this relative stability, life satisfaction is still responsive to changes in circumstances such as receiving therapy or illness of spouse (Pavot and Diener, 1993a). Analysis of large datasets using within person change

---

\(^{54}\) Studies on monozygotic and dizygotic twins lead to the conclusion that 55% of the variance in negative emotionality, and 40% of the variance in positive emotionality was accounted for by genetics, and that in the long run up to 80% of SWB is heritable (Lykken and Tellegen, 1996).
identifies statistically significant relationships, such as marriage and employment status, with single life satisfaction questions, even using only a 3-point scale.

In terms of evaluating projects and local policies reliance on very large datasets is problematic. Weighted-domain satisfaction has the potential for greater sensitivity, and the advantage of individual domain satisfaction information which may potentially be more sensitive to a particular policy change.

Scales derived from a combination of well-being accounts or focusing on mental health (the Affectometer 2, the GHQ, and the CES-D), have many more items and therefore a greater possible range across the full scale.

*Summary of sensitivity*

There will be little benefit for policy in a highly sensitive yet non-valid measure of well-being. Hence it is necessary to have a very clear idea of the appropriate concept then find ways of making this more sensitive. The discriminant ability of individuals when assessing their lives may be limited, such that it is not possible to generate more sensitive subjective outcome measures. However, if it is known from analysis of large datasets that certain factors enhance SWB, it may be possible to use those factors as intermediate outcome measures for assessing policy.

### 5.4.4 Practical and efficient to collect

*Preference satisfaction*

Accurate measures of income and consumption are expensive and time consuming to collect, and are subject to misreporting and low response rates. Detailed expenditure diaries are required for accurate measures of consumption and accurate measures of income require cross checking against all possible income sources. However, less detailed measures of consumption and income are easily obtainable as simple self-report.

Furthermore, income may already be collected for other purposes, such as monitoring taxation revenues. This may, however, be less relevant for those groups at risk of low well-being, such as single mothers and the elderly, who obtain income from a range of different sources including private and government transfer income and off-the-books income (Meyer and Sullivan, 2003).

The costs of obtaining accurate income data depend upon the extent of self-employment within the population and attitudes towards revealing income details to researchers. Income
data is likely to be considerably cheaper to collect than consumption data, for example, “In the U.S., it costs five times as much per household to collect consumption (and other) information in the Consumer Expenditure Survey (CEX) as it does to collect income (and other) data in the Current Population Survey (CPS)” (Deaton and Zaidi, 2002: 13).

The practicality of collecting health-related utility measures depends upon the method used. HRQoL may be assessed directly (e.g. using TTO or SG on the individual’s current health state), however, this usually involves reasonably lengthy interviews. Alternatively, the use of validated health description instruments that can be linked to existing valuation studies enables the use of brief self-report questionnaires.

Subjective: Hedonist

Of the scales discussed which track positive and negative affect the ESM is likely to prove prohibitively expensive for most circumstances. The DRM is less expensive and time consuming (although still takes about one hour), and future developments of the DRM may make this even less time consuming. As a potentially sensitivity measure, the DRM may enable sample sizes to be small, so should not be overlooked on practicality grounds for all circumstances.

Subjective: Evaluative

Single life satisfaction questions have an obvious advantage in terms of time and survey space and have been included in many large surveys, and found to have high response rates. Similarly, domain satisfaction responses and multiple item evaluative measures (such as SWLS) are quick and easy to complete.

Summary of practicality

Whether the measure is practical depends on its intended purpose. For research specifically targeted at understanding well-being, time and resource constraints will be less pressing. Evaluation of specific policies is more likely to require survey instruments that can be completed fairly quickly and by self report. Some situations, such as health care trials, will have limited sample sizes due to limits in the relevant subgroup and as such require more sensitive measures. Similarly, where the impact on any single individual is expected to be small (such as the improvement of an environmental feature) sensitivity will be essential even if sample sizes are considerable.
5.4.5 Public/Political acceptability

Preference satisfaction

Well-being as the satisfaction of individual preferences may be attractive to the public in that the individual judges what will be in their best interests. Whilst individuals may make mistakes in their judgements, as individuals we may see mistakes in others yet still have confidence in our own preferences. Even if preference satisfaction could have public support, the means by which it is measured, namely revealed preferences within the market, or WTP for non-market goods, is less likely to be consistent with public views. An individual may have complete confidence that their preferences reflect their best interests without full opportunity to express those preferences.

The maximisation of income without constraints arising from direct concern over health care, education, safety, the environment etc. would risk alienating public support. However, there is general acceptance towards a government role in enhancing average incomes, and little resistance to the use of income as a measure of value within economic evaluation, or as input into the theoretical support for economic policy. As discussed in Chapter One, evidence on the failure of rising incomes to bring improvement to subjective assessments of peoples’ lives may begin to undermine public support for political action aimed at maximising economic growth.

Subjective: Hedonist

Public support for hedonist measures face an additional problem due to common use of language, specifically the various meanings of the term ‘happiness’. Whilst the term ‘happiness’ may be perceived as trivial (in part because it can refer to temporary positive feelings, rather than a reflection of how well life is going) the term ‘well-being’ is far less likely to be perceived as trivial. Where hedonist measures are picking up the totality of mental states rather than feelings of happiness and sadness this should have greater public legitimacy.

Subjective: Evaluative and other SWB

At the measurement level, evaluative measures of well-being appear the least paternalistic in the sense that the individual alone determines their well-being. However, at the policy level, paternalism reappears in the sense that policy is based on maximising reported well-being rather maximising individual choice. This has led to objections, for example;

“the logical conclusion of much happiness research – that individuals’ own judgements about what is good for them can be overridden by experts wielding clipboards and
However, this type of criticism is about how the measures are used, rather than the measures themselves. Policies targeted at maximising income could be equally criticised as being contrary to ‘individual’s own judgements about what is good for them’.

Concerns may surround the potential of subjective measures to be manipulated by governments and/or respondents. For example, McMahon argues that a gross national happiness indicator would be “especially susceptible to government manipulation and citizen distortion” (McMahon, 2007). Once the public become aware of a political role for SWB measures they may be open to deliberate attempts to reflect individual opinion and support for current policies. This may be most acute in domain satisfactions, for example, satisfaction with health may be reported as lower where individuals wish to see a change with the funding or operation of centrally provided health care.

Manipulation on behalf of governments may be restricted by the transparency of national statistics organisations, however, whilst there are many potential measures which arguably reflect well-being choice of measure is likely to be influenced by those which most closely reflect the story desired by policy makers.

*Summary of compatibility with public views*

Concerns over lack of public support may help point to cases where the measure lacks face validity. However, since public attitudes are dynamic they should not override concerns over validity.

**5.5 Conclusion**

This assessment has shown that all the well-being accounts and instruments discussed face reasonable criticism and none fully meets all the criteria established in Chapter Four.

The most important of all the criteria is validity, since lack of validity may lead to policy which enhances something other than well-being – which may or may not have value.

Although income does tell us something about how well an individual’s life is going, it is clearly not a valid measure of well-being. WTP provides a measure of trade-off - how much the individual at that moment in time thinks they would be willing to trade-off alternative uses of money, but it is not an accurate measure of their actual change in well-being following acquisition of the good or service.
The validity of hedonist accounts of well-being will always be in doubt because of intuitively held beliefs that positive and negative feelings are not the only thing which determines how well a life goes for the person living it. Evaluative accounts have the potential to address the limitations of measures based only on positive and negative feelings. Current evidence suggests that SWB have reasonable validity.

Evaluative measures adhere to the criteria equally as well if not better than other measures of well-being. This would lead us to expect them to be useful in a policy context, and indeed perform better than existing measures. This will be explored over the next three chapters.

These chapters will consider how useful subjective measures are currently for policy, in terms of national and local policy and conducting economic evaluation. The first will consider policy implications arising from our understanding of the relationship between income and SWB. Chapter Seven will investigate a potential area where analysis of SWB has led to policy recommendations that would not arise from the maximisation of income or consumption: namely increased community involvement and neighbourhood interaction. Chapter Eight will evaluate the use of SWB as a means of valuing non-market goods.
### APPENDIX 5.A

#### Table 20: Estimates of temporal stability of SWB measures

<table>
<thead>
<tr>
<th>Study</th>
<th>Measure (response scale)</th>
<th>Test-retest</th>
<th>Temporal interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multi-item</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alfonso et al. (1996)</td>
<td>SWLS</td>
<td>0.83</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Pavot et al. (1991)</td>
<td>SWLS</td>
<td>0.84</td>
<td>1 month</td>
</tr>
<tr>
<td>Eid and Diener (2004)</td>
<td>SWLS</td>
<td>0.74</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Blais et al., (1989)a</td>
<td>SWLS</td>
<td>0.64</td>
<td>2 months</td>
</tr>
<tr>
<td>Diener et al., (1985)</td>
<td>SWLS</td>
<td>0.82</td>
<td>2 months</td>
</tr>
<tr>
<td>Yardley &amp; Rice (1991)</td>
<td>SWLS</td>
<td>0.50</td>
<td>10 weeks</td>
</tr>
<tr>
<td>Schimmack et al., (2002)</td>
<td>SWLS</td>
<td>0.74</td>
<td>3 months</td>
</tr>
<tr>
<td>Lepper (1998)</td>
<td>SWLS</td>
<td>0.75</td>
<td>9 months</td>
</tr>
<tr>
<td>Magnus et al., (1993)a</td>
<td>SWLS</td>
<td>0.54 (0.52</td>
<td>4 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>when</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>informant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>reports in</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>second period</td>
<td></td>
</tr>
<tr>
<td><strong>Single-item</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andrews &amp; Whithey (1976)</td>
<td>Life satisfaction (Terrible 1 – Delighted 7)</td>
<td>0.64</td>
<td>1 hour</td>
</tr>
<tr>
<td>Glatzer (1984)b</td>
<td>Life satisfaction</td>
<td>0.6</td>
<td>1 hour</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Life satisfaction (1-4)</td>
<td>0.59</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Pavot &amp; Diener (1993b)</td>
<td>Life satisfaction (1-7)</td>
<td>0.41</td>
<td>1 month</td>
</tr>
<tr>
<td>Pavot &amp; Diener (1993b)</td>
<td>Overall happiness (1-7)</td>
<td>0.71</td>
<td>1 month</td>
</tr>
<tr>
<td>Chamberlain &amp; Zika (1992)</td>
<td>Life satisfaction (mothers) ('Life 3': Terrible 1 - Delighted 7, average of questions asked twice).</td>
<td>0.62</td>
<td>3 months</td>
</tr>
<tr>
<td>Chamberlain &amp; Zika (1992)</td>
<td>Life satisfaction (aged 60+)</td>
<td>0.68</td>
<td>3 months</td>
</tr>
<tr>
<td>Ehrhardt et al. (2000)</td>
<td>Life satisfaction (0-10)</td>
<td>0.45-0.58</td>
<td>1 year</td>
</tr>
<tr>
<td>Atkinson (1982)</td>
<td>Life satisfaction (1-11)</td>
<td>0.41</td>
<td>2 years</td>
</tr>
<tr>
<td>Horley and Lavery (1991)</td>
<td>Life satisfaction (0-10)</td>
<td>0.25</td>
<td>7 years</td>
</tr>
<tr>
<td>Ehrhardt et al. (2000)</td>
<td>Life satisfaction (0-10)</td>
<td>0.29</td>
<td>10 years</td>
</tr>
<tr>
<td>Fujita and Diener (2006)</td>
<td>Life satisfaction (0-10)</td>
<td>0.51</td>
<td>Av. of 1st 5 and last 5 yrs (17 yrs of data)</td>
</tr>
<tr>
<td>Atkinson (1982)</td>
<td>Domain weighted (5 domains)</td>
<td>0.58</td>
<td>2 years</td>
</tr>
<tr>
<td>Smith (1979)</td>
<td>Happiness (1-3)</td>
<td>0.6</td>
<td>1 month</td>
</tr>
<tr>
<td>Lepper (1989)</td>
<td>General happiness (1-7)</td>
<td>0.53</td>
<td>9 months</td>
</tr>
<tr>
<td>Atkinson (1982)</td>
<td>Happiness (1-3)</td>
<td>0.39</td>
<td>2 years</td>
</tr>
<tr>
<td>Palmore and Kivett (1977)</td>
<td>Cantrils ladder</td>
<td>0.70</td>
<td>2 years</td>
</tr>
<tr>
<td>Headey and Wearing (1989)</td>
<td>SWB</td>
<td>0.5-0.6</td>
<td>6 years</td>
</tr>
<tr>
<td>Dennerstein, Lehert and Guthrie's (2002)</td>
<td>Affectometer 2</td>
<td>0.60</td>
<td>3 time points</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Self-acceptance (PWB)</td>
<td>0.85</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Positive relations with others (PWB)</td>
<td>0.83</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Autonomy (PWB)</td>
<td>0.88</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Environmental mastery (PWB)</td>
<td>0.81</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Purpose in life (PWB)</td>
<td>0.82</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Authors</td>
<td>Method</td>
<td>Correlation</td>
<td>Timeframe</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Ryff (1989)</td>
<td>Personal growth (PWB)</td>
<td>0.81</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Chamberlain &amp; Zika (1992)</td>
<td>Psychological well-being</td>
<td>0.72-0.87</td>
<td>3 months</td>
</tr>
<tr>
<td>Reker &amp; Wong (1984)</td>
<td>Psychological well-being</td>
<td>0.79</td>
<td>2 years</td>
</tr>
<tr>
<td></td>
<td>(Perceived well-being scale)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive affect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradburn (1969)</td>
<td>Positive affect (ABS)</td>
<td>0.83</td>
<td>3 days</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Positive affect</td>
<td>0.68</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Watson et al. (1988)</td>
<td>Positive affect (PANAS)</td>
<td>0.58</td>
<td>2 months</td>
</tr>
<tr>
<td>Chamberlain &amp; Zika (1992)</td>
<td>Positive affect</td>
<td>0.63</td>
<td>3 months</td>
</tr>
<tr>
<td></td>
<td>(Affectometer) (mothers)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepper (1998)</td>
<td>Positive affect</td>
<td>0.49</td>
<td>9 months</td>
</tr>
<tr>
<td>Negative affect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradburn (1969)</td>
<td>Negative affect (ABS)</td>
<td>0.81</td>
<td>3 days</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Negative affect</td>
<td>0.60</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Watson et al. (1988)</td>
<td>Negative affect (PANAS)</td>
<td>0.48</td>
<td>2 months</td>
</tr>
<tr>
<td>Chamberlain &amp; Zika (1992)</td>
<td>Negative affect</td>
<td>0.56</td>
<td>3 months</td>
</tr>
<tr>
<td></td>
<td>(Affectometer) (mothers)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepper (1998)</td>
<td>Negative affect</td>
<td>0.52</td>
<td>9 months</td>
</tr>
<tr>
<td>Affect balance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steptoe et al. (2005)</td>
<td>ESM</td>
<td>0.65</td>
<td>Weekend to weekday</td>
</tr>
<tr>
<td>Bradburn (1969)</td>
<td>Affect balance (ABS)</td>
<td>0.76</td>
<td>3 days</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Net affect (DRM)</td>
<td>0.64</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Difmax (DRM)</td>
<td>0.60</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Krueger &amp; Schkade (2006)</td>
<td>Uindex (DRM)</td>
<td>0.50</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Lepper (1998)</td>
<td>Affect balance</td>
<td>0.55</td>
<td>9 months</td>
</tr>
<tr>
<td>Schimmack and Oishi (2005)</td>
<td>Meta-analysis on 83 coefficients on 38 independent samples on life satisfaction</td>
<td>Range [0.24 to 0.87]</td>
<td>Various</td>
</tr>
</tbody>
</table>

*a* cited in Pavot and Diener (1993)  
*b* cited in Schimmack and Oishi (2005)
Chapter 6: Income and subjective well-being: Policy implications

6.1 Introduction

Chapter Five showed that subjective measures fare at least as well as other measures of well-being according to a set of desirable criteria, and come closer to meeting validity conditions than the more traditional measures of income and consumption. This suggests that subjective measures may have a great deal to offer policy makers.

Perhaps the most important contribution of the SWB literature to date is the questioning of the relationship between income and well-being. This chapter considers three areas where new understanding of the relationship between income and SWB points to a role for government intervention. For each it assesses the evidence derived from current literature and supplements this with additional analysis, predominantly to gain a better understanding of the processes involved in these relationships. The three policy relevant claims drawn from the literature considered here are:

1. Unforeseen adaptation to income may result individual failure to maximise SWB (Layard, 2005b)
2. Individuals may collectively overwork because they do not account for the negative externality their work imposes upon others (Frank, 1999; Layard 2005b).
3. Individuals may collectively over-consume because relative consumption exacerbates self-control problems (Frank, 1999).

Each of these claims will be outlined and critically reflected upon in the light of recent evidence and additional analysis from the BHPS. This will involve a set of new hypothesis to be tested for each claim. Policy implications will be discussed in the concluding section.

6.2 Unforeseen adaptation to income

Individuals may fail to maximise SWB in the long run if they do not consider the impact of current income upon future SWB. Adaptation theory (Helson, 1964) predicts that individuals become accustomed to new levels of income, and consequently, reaching new levels of income only impacts upon SWB in the short run. If individuals make decisions without accounting for
adaptation they are likely to make inefficient choices in terms of the trade-off between income and leisure. This is demonstrated in detail in Appendix 6.A.

Adaptation to income arises partly through adaptation to stimuli, and partly through raising expectations. Income expectations are likely to be determined by a range of sources including past income, the income of others and expectations of salary increases (Koszegi and Rabin, 2006).

If higher past income raises expectations and desires, then controlling for current income, we would expect past income to be negatively related to SWB. Furthermore, if adaptation to new levels of income is complete, actual income levels would not be expected to impact upon SWB in the long run, but changes in income levels may have a short-term effect. When changes in income over time are not controlled for, a relationship between income and SWB may be found if those on higher incomes have experienced more positive changes than those on lower incomes (Clark et al., 2007). In the long term, if individuals fully adapt to new levels of income, we would expect SWB to be positively related to growth rates, but not income levels.

These issues have been explored using survey data which considers the impact of past income on job satisfaction (Clark, 1999; Clark et al., 1998; Grund and Sliwka, 2003), income satisfaction (Burchardt, 2004) and satisfaction with life (Di Tella et al., 2007).

6.2.1 Current evidence on adaptation to income and SWB

There is reasonable evidence that higher incomes raise income expectations and, controlling for current income, having higher income expectations is detrimental to life satisfaction.

Aspirations have been found to increase as income rises. For example, Stutzer (2004) estimates that for Switzerland the income perceived to be sufficient (which is interpreted as aspiration income) increases by 4.2% following a 10% increase in income. Similarly, Easterlin (2005b) found that material aspirations, in terms of the number of ‘big ticket’ consumer items perceived to make up the ‘good life’, increase in line with income.

A rise in income may have a cost in terms of raising expectations, potentially above that which can be met by current income. For example, Graham and Pettinato (2001) analyse Latinobarometro data (1997-2000) and identify a group of ‘frustrated achievers’ who are unhappy despite a rapid growth in income because of rising aspirations.

The Leyden school has conducted a body of research showing that reported verbal labels (e.g. “good” “bad” or “sufficient”) used to describe levels of income and the income level given to
match each verbal label increases as income increases, a concept that van Praag named ‘preference drift’ (van Praag, 1971). Estimates of ‘preference drift’ have been taken from the coefficient of own logged income against the log of the minimum income question (van Praag and Kapteyn, 1973). This coefficient falls between zero to one, where zero implies current income does not influence judgments of reported minimum income and one suggests the judgment is entirely relative to own income. In some models family size is controlled for along with personal and geographical characteristics. Estimates of preference drift range from 0.14 to 0.54 depending upon the choice of model and dataset (Milanovic and Jovanovic, 2003; Garner and Short, 2005), suggesting considerable adaptation to new levels of income.

However, own income is likely to act as reference point for responses, creating a framing effect. Furthermore, whilst suggestive of adaptation, these studies do not indicate the duration of past income effects, nor the impact of higher expectations on well-being. The consequence of income expectations on SWB has been explored through analysis that uses SWB as the dependent variable.

Using US data MacDonald and Douthitt (1992) find that higher expected income decreases a domain-satisfaction index and the difference between own income and the level of income judged as ‘terrible’ has a positive effect.

Various studies have shown the importance of annual changes in income. Inglehart and Rabier (1986) find life satisfaction and happiness to be unrelated to current income, but positively correlated with a change in financial position in Eurobarometer data (1973-1983). Change in income has been found to be related to perceived income adequacy in Singapore and Taiwan (Chan et al., 2002), and economic welfare in Russia (Ravallion and Lokshin, 2001).

Analysis of the 1995 and 2000 waves of the BHPS (Brown et al., 2005) finds a strong association between the perception of financial deterioration since the previous year and reduced psychological well-being, also highlighting the importance of changes in income.

If change matters, and past incomes raise expectations, we would expect that, controlling for current income, past income would have a negative impact on SWB. Studies incorporating lagged income have tended to find a negative but only weakly significant effect. Di Tella et al. (2007) analyse West German GSOEP data (1984-2000) to find negative coefficients for the previous four years of income, with only about 40% of the initial income effect remaining after four years. Although none of the lagged income coefficients are individually significant, F tests show that jointly there is a significant amount of adaptation after the current year. They also

---

55 Measured by the GHQ Caseness converted into a dichotomous variable.
find that average income over the last five years is not significantly related to life satisfaction (except for those with right wing political views), which supports the presence of income adaptation. Weinzierl (2005) also uses the West German GSOEP data but controls for relative income in addition to own income (using equivalised three year moving average income) and finds a negative, yet insignificant effect of lagged (one period) log income.

However, Graham et al. (2004) find that income five years earlier increases life satisfaction when controlling for current income in the Russia Longitudinal Monitoring Survey (RLMS, 1995 & 2000). This suggests that the relationship between past income and SWB is indeterminate. It may reduce SWB if it leads to unfulfilled expectations, but may also proxy for assets and previous wealth accumulation that can increase SWB. However, the differential impact of past income noted in the studies here may reflect differences between high and medium income countries. Furthermore, the income-SWB relationship in Russian data has been shown to be at odds with European findings (Senik, 2004).

One of the difficulties with exploring adaptation effects is that little is known of the time path in which adaptation operates or the formation of income expectations. Influential income may be last year, five years ago or even as far back as one’s childhood. McBride’s analysis of the 1994 GSS finds that, controlling for current income, having a standard of living that is much worse than one’s parents at the same age has a negative effect on overall happiness. Indeed, controlling for this, own income is non-significant (McBride, 2001). Parental income is likely to feed into income expectations and, controlling for income, those with higher expectations are likely to judge their current income less favourably which may consequently lower SWB.

6.2.2 New hypothesis to aid the understanding of adaptation to income and SWB

Although the evidence points to a negative impact of past income on SWB, this finding arises from few studies with limited significance. If this finding is robust it should be detectable in a range of different datasets. It is therefore useful to consider whether any impact of adaptation to income can be identified within the BHPS. The first hypothesis to be tested therefore is simply the presence of negative lagged income coefficients.

Hypothesis i: Controlling for current income, lagged income will have a negative relationship with life satisfaction, against a null of non-significance or positive effect.

56 A similar effect is also found by Kingdon and Knight (2007) in South Africa.

57 This should not be interpreted as evidence that high expectations in themselves are bad, since they may be part of a dynamic process by which those with higher expectations are motivated to gain higher incomes.
Before any potential policy implications can be drawn from evidence of adaptation it is necessary to understand more fully what is causing this relationship. The BHPS can be used to explore some alternative explanations.

If a household’s medium term expenditure commitments are a function of their current income, then these commitments will be carried forward independently of future income. Adaptation to new levels of income in this context arises when consumption commitments adjust to the new level of income, implying that consumption decisions follow a partial adjustment model.

If reduced satisfaction with life arises because expenditure commitments are out of line with income, we would expect those with more commitments to experience a stronger negative effect of past income. Since mortgages are a good example of a fairly inflexible expenditure commitment, those with a mortgage would be expected to experience a stronger relationship with past income.

Hypothesis ii: Controlling for current income, past income is more strongly related to life satisfaction for those with a mortgage, against a null of no difference or lower effect.

Past income levels may also determine norms of expenditure behaviour, with higher past incomes resulting in more expensive desires. The presence of both expenditure commitments and expectations and the possible frustration of material desires would be expected to be particularly influential on income satisfaction and perceptions of the sufficiency of income.

Income satisfaction is one of the domain questions asked in the BHPS, responded to on a 1 to 7 scale. Perceptions of the sufficiency of income (financial coping) are surveyed by the question: “How well would you say you yourself are managing financially these days? Would you say you are... living comfortably, doing alright, just about getting by, finding it quite difficult, finding it very difficult”. These responses are converted into a 1 to 5 scale, where 5 represents living comfortably.  

If expenditure habits and desires do not respond immediately to changes in income it would be expected that for any given level of current income, lagged income should have a negative impact upon income satisfaction. Likewise where past incomes were low, desires should be more easily met with new higher levels of income.

Hypothesis iii: Controlling for current income, lagged income is negatively related to satisfaction with income and perceptions of income adequacy, against a null of non-significance or positive effect.

58 This variable is ordinal, yet for consistency, and ease of controlling for individual effects, it is treated cardinally. The main result is the same when it is analysed adhering to the ordinal nature of the data using a random effects ordered probit, although not controlling for individual heterogeneity.
If the negative effect of past income is driven by financial commitments it would also be expected that this would be felt only in declines in income. If the impact of lagged income is considered separately for those who have experienced a rise in income and those who have experienced a decline, the negative impact of lagged income would be expected to be stronger for the latter.

**Hypothesis iv:** Those who have experienced a decline in income since the previous period will show a greater negative impact of the previous period income, against a null of no difference or less negative impact.

Another possibility is that the negative impact of past income arises from the pleasure derived from feeling positive about the future. Evidence of a positive change in income or upward trend may be taken as a sign that things are improving. Increases in income may cause positive assessments about the future and the direction in which life is heading. Likewise decreases in income may result in pessimism about the future. Hence the negative coefficient on past income may not only be about adaptation but how individuals see their future progressing.

For example, Burchardt (2004) finds that those with constant income are less happy with their income than those with rising income, implying either adaptation to income or a benefit arising from positive expectations. The benefit of rising trajectories is also found in Russian data by Senik (2006).

The BHPS includes a question which asks: “*Looking ahead, how do you think you will be financially a year from now, will you be better than now, worse than now, same as now?*” Controlling for responses to this question will enable a distinction to be made between adaptation and the implication of changing incomes for future outlook and the expectations of future finances.

**Hypothesis v:** Past incomes will no longer be significant when controlling for financial expectations, against a null that the coefficient on past income will be robust to the inclusion of financial expectations.

An alternative explanation for the detrimental impact of past income is that higher income crowds out other well-being generating attributes, and this is experienced with some delay. For example, higher income may result in reduced contact with friends and family. The negative relationship between income and social contact was discussed in Chapter Three. Although having high levels of contact with friends, family and neighbours is controlled for within these models this is unlikely to be sufficient to control for the quality of these relationships. If relationships are damaged this may have future consequences in terms of reduced SWB.

Predictions of SWB, particularly when controlling for fixed effects, explain only a very small
amount of the variance within SWB. Hence a large component remains unaccounted for, opening the possibility of omitted variables with a negative relationship to life satisfaction and positive relationship to past incomes.

One such variable may be materialist attitudes and consumerism, which have been linked to low psychological well-being (Kasser, 2002). Increased income may result in increased materialist values, and less pro-social behaviour. The relative importance given to income, and whether this changes following an increase in income, can be explored within the BHPS. Waves 8 and 13 contain a question on the importance of health, money, having children, having a good job, being independent, owning one’s own home, good partnership, and good friends, valued on a scale of 1 to 10.\(^\text{59}\) Simply using the response to the importance of income is problematic since an individual may respond that all items are extremely important. The relative importance of income is therefore used. This is the ratio of importance of income to the average importance of good partnership, good friends and good health. Testing whether the importance of income offers an explanation for the negative impact of past income involves two stages, firstly showing that changes in income relate to changes in the importance of income, and secondly that changes in the importance of income are negatively related to life satisfaction.

\textit{Hypothesis vi: Income is positively related to the importance of income, against a null of non-significance or negative impact.}  

\textit{Hypothesis vii: The importance of income is negatively related to life satisfaction against a null of non-significance or positive impact.}

\subsection*{6.2.3 Testing new hypothesis on the BHPS}

The first hypothesis requires confirming negative past income effects in another panel dataset. In line with Di Tella et al. (2007) the impact of income lagged for four periods upon life satisfaction was considered using OLS fixed effects on data from the BHPS.\(^\text{60}\) The BHPS analysis supports the Di Tella and colleagues finding of a negative lagged income effect.

The income measure used is logged, household, net, current income. Since household rather than individual income is used standard errors are adjusted for clustering at the household level. Standard controls include: Education (no qualifications compared with degree, A level, O level, commercial or CSE); Marriage status (married compared with divorced or separated,

\textsuperscript{59} The questions reads: “I’m going to read you a list of things that people value. For each one I’d like you to tell me on a scale from 1 to 10 how important each one is to you, where ’1’ equals ‘Not important at all’, and ’10’ equals ’Very important’.” \textsuperscript{60} Chapter Three gives an outline of this method and justification for its use with SWB data. In this, and other analysis in this chapter the sample excludes those still studying, those still at school and those under 18, and those living in Northern Ireland. Wave 15 is not included since net income was not available for this wave.
cohabiting, widowed, never married); Health (self-rated health good compared with poor or very poor, fair, or excellent; reports a problem walking hospital stay in the last 12 months, reports to be disabled); Age (in 5 year categories); Social contact (dummy for seeing friends or family most days or weekly, talking to neighbours most days or weekly); Job status (employed compared to long term sick, retired, unemployed, self-employed, family carer, government training, or other employment status); Caring role (dummy for carers for another person over 50 hours a week); Household composition (number of adults, number of adults squared, number of children 0-4, 5-11 and 12-18); Urbanisation (dummy for metropolitan regions); wave dummies; and regional dummies. Full details of all the variables used in these regressions can be found in the appendix for Chapter Three (3.D).

Table 21: OLS-FE on life satisfaction, including past income

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income</td>
<td>.0533(.0137)**</td>
<td>.0471(.0183)**</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>-.0320(.0163)*</td>
<td>-.0320(.0163)<em>[</em>]</td>
</tr>
<tr>
<td>Log hh income : t-2</td>
<td>-.0320(.0171)</td>
<td>-.0184(.0159)</td>
</tr>
<tr>
<td>Log hh income : t-3</td>
<td>-.0184(.0159)</td>
<td>-.0269(.0162)[*]</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0404</td>
<td>0.0423</td>
</tr>
<tr>
<td>Observations</td>
<td>64,056</td>
<td>38,837</td>
</tr>
<tr>
<td>Individuals</td>
<td>15,258</td>
<td>9,631</td>
</tr>
</tbody>
</table>

Note: Waves 7-10&12-14. * P < 0.05, ** P< 0.01, throughout, [*] denotes significance level under a one tailed test where appropriate for the hypothesis. Coefficient(se)

As can be seen in Table 21 the current income effect is positive but fairly weak and changes little when the lagged income variables are included. The strongest negative impact comes from income in the previous year, which is the only individually significant lag, and is about two-thirds the size of the current income coefficient. Hence the null from the first hypothesis can be rejected. An F test does not reject the null that other three lagged terms are jointly non-significant at (F(3, 24503)=1.55, Prob>F=0.2006). In this model therefore the effect is only significant for the first lag.

The second hypothesis considers the effect of lagged income on those with and without mortgages, as a proxy for expenditure commitments. Table 22 shows that it is not the case that those with mortgages suffer more from higher past incomes⁶¹, indeed the opposite is the case. Hence it is not possible to reject the null for the second hypothesis. It may be that the presence of a mortgage is a poor indicator of financial commitments faced by the household. Mortgages

---

⁶¹ Supported by a Wald test.
may also indicate higher wealth accumulated from past income from which the individual benefits in the current period.

Table 22: OLS-FE on life satisfaction, including past income, for those with and without a mortgage

<table>
<thead>
<tr>
<th></th>
<th>With a mortgage</th>
<th>Without a mortgage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
<td>.0740(.0278)**</td>
<td>.0346(.0260)</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>-.0003(.0243)</td>
<td>-.0547(.0232)*</td>
</tr>
<tr>
<td>: t-2</td>
<td>.0140(.0263)</td>
<td>-.0081(.0239)</td>
</tr>
<tr>
<td>: t-3</td>
<td>-.0150(.0250)</td>
<td>-.0270(.0222)</td>
</tr>
<tr>
<td>: t-4</td>
<td>-.0247(.0255)</td>
<td>-.0369(.0225)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0430</td>
<td>0.0489</td>
</tr>
<tr>
<td>Observations</td>
<td>18,616</td>
<td>20,221</td>
</tr>
<tr>
<td>Individuals</td>
<td>5,103</td>
<td>5,763</td>
</tr>
</tbody>
</table>

Note: waves 7-10&12-14. Coefficient(se).

The third hypothesis explores the impact of past income upon income satisfaction and ability to cope financially. Table 23 shows that, against expectations, past income has a positive impact on both income satisfaction and financial coping. It is not possible to reject the null for the third hypothesis.

Table 23: OLS-FE on income satisfaction and perceptions of financial coping

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: Income satisfaction</th>
<th>Dependent variable: Financial coping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
<td>.4162(.0253)**</td>
<td>.2007(.0153)**</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>.0872(.0229)**</td>
<td>.0380(.0138)**</td>
</tr>
<tr>
<td>: t-2</td>
<td>.0525(.0229)*</td>
<td>.0219(.0135)</td>
</tr>
<tr>
<td>: t-3</td>
<td>.0115(.0220)</td>
<td>-.0093(.0135)</td>
</tr>
<tr>
<td>: t-4</td>
<td>-.0336(.0210)</td>
<td>.0091(.0131)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0470</td>
<td>0.0369</td>
</tr>
<tr>
<td>Observations</td>
<td>38,981</td>
<td>34,472</td>
</tr>
<tr>
<td>Individuals</td>
<td>9,633</td>
<td>8,946</td>
</tr>
</tbody>
</table>

Note: Waves 7-14, excluding wave 11 on income satisfaction. Coefficient(se).

The positive impact of past income upon income satisfaction and financial coping may imply a role for wealth accumulation. For example, higher previous incomes may reduce current expenditure needs through lower monthly mortgage repayments.

Burchardt (2004) analyses waves six to ten of the BHPS and also finds that a positive change in household income is negatively related to satisfaction with income when controlling for household income and a range of household and personal characteristics including fixed effects. However, analysed by quintiles, she finds that this negative effect is only significant for the bottom income quintile, and becomes significantly positive for the top income quintile. For
those in the bottom quintile, recently reaching a certain level of income may suggest previous years of low income, hence the possible presence of debts and low levels of assets. For the top income quintile the positive effect of change, combined with a positive effect of income level, suggests that changes matter in addition to levels. She also finds that experiencing both falls in income and large increases in income reduce income satisfaction compared to experiencing constant income, which is a surprising result without an obvious explanation.

The fourth hypothesis proposes that those who have experienced a recent fall in income will show a greater past income effect. When data are divided according to those who experienced a loss since the previous period and those who did not (Table 24) those with decreases in income do show a greater loss in life satisfaction from the past income level. However, when the data are split in this fashion (and controls for additional lags on income are not included) neither current income nor lagged income remain significant, highlighting a lack of robustness of the income and lagged income effect.

Table 24: OLS-FE on life satisfaction, for those with rising and falling incomes

<table>
<thead>
<tr>
<th>Last period income higher than current income</th>
<th>Last period income lower than current income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
<td>.0410(.0304)</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>-.0258(.0331)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0465</td>
</tr>
<tr>
<td>Observations</td>
<td>25,047</td>
</tr>
<tr>
<td>Individuals</td>
<td>11,262</td>
</tr>
</tbody>
</table>

Note: Waves 7-10&12-14. Coefficient(se).

The fifth hypothesis proposes that part of the past income effect arises due to the interpretation of income trends on future expectations. The inclusion of a future expectations variable brings little change to the size of the income or lagged income coefficients (Table 25). The significance of the pessimism variable highlights the importance of expectations and concerns for the future, however, it does not suggest that the impact of lagged income is due to expectations of a financial trend continuing.

Table 25: OLS-FE on life satisfaction, including future expectations

<table>
<thead>
<tr>
<th>Without future expectations (same sample)</th>
<th>Including future expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
<td>.0353(.0185)</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>-.0289(.0168)</td>
</tr>
<tr>
<td>Log hh income : t-2</td>
<td>-.0009(.0175)</td>
</tr>
<tr>
<td>Log hh income : t-3</td>
<td>-.0129(.0164)</td>
</tr>
</tbody>
</table>

A Wald test does not reject a null hypothesis that these two coefficients are equal.
Chapter 6  Income and SWB: policy implications

The six and seventh hypotheses explore the relationship between income and judgements on the importance of income, and importance of income and SWB. Placing more importance on income compared to health, intimate relationships and friends, is negatively correlated with life satisfaction (-0.08). When controlling for other determinants of life satisfaction, it has a significantly negative relationship within a random effects model but the coefficient is not significant in a fixed effects model. Given that this data is only available in two waves non-significance within the fixed effects model is not surprising. However, since the relative importance of income is likely to be related to personality, and hence the individual unobserved effect, it is not possible to draw any strong conclusion from these results.

Table 26: GLS-RE and OLS-FE on life satisfaction, including the importance of income

<table>
<thead>
<tr>
<th></th>
<th>RE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
<td>.0550(.0316)</td>
<td>.0593(.0673)</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
<td>.0098(.0333)</td>
<td>-.0287(.0676)</td>
</tr>
<tr>
<td>Log hh income : t-2</td>
<td>-.0094(.0345)</td>
<td>.0111(.0659)</td>
</tr>
<tr>
<td>Log hh income : t-3</td>
<td>.0228(.0355)</td>
<td>-.0355(.0689)</td>
</tr>
<tr>
<td>Log hh income : t-4</td>
<td>.0173(.0290)</td>
<td>-.0431(.0610)</td>
</tr>
<tr>
<td>Relative importance of income</td>
<td>-.2265(.0548)<strong>[</strong>]</td>
<td>.0743(.1365)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>Overall 0.1937</td>
<td>Within 0.0593</td>
</tr>
<tr>
<td>Observations</td>
<td>11,830</td>
<td>11,830</td>
</tr>
<tr>
<td>Individuals</td>
<td>8,393</td>
<td>8,393</td>
</tr>
</tbody>
</table>

Note: Waves 8&13. Coefficient(se).

The determinants of the importance of income can be explored by treating this as the dependent variable. As Table 27 shows, using the standard set of controls, a higher household income is positively related to the relative importance of income (this uses a random effects model as no variables are significant within the fixed effects model), but is only weakly significant (p value of 0.064). However, the lagged income coefficient is not significant, therefore it is not possible to say whether higher incomes of the past leave a legacy of attitude towards income.

---

63 Although significant at the 5% level using the one tailed test in line with the hypothesis.
The perceived relative importance of income is lower for those in excellent health, women, the self-employed, and for the earlier wave. Interestingly, being divorced or never married compared to being married strongly increases the relative importance of income. This brings into question the direction of causality - does success in personal relationships lead them to be judged as more important? If so, the negative impact of the relative importance of income shown in Table 26 may arise from current difficulties in personal relationships rather than lower priority being given to relationships.

**Table 27: GLS-RE on the relative importance of income**

<table>
<thead>
<tr>
<th>Coefficient(se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hh income : t</td>
</tr>
<tr>
<td>Log hh income : t-1</td>
</tr>
<tr>
<td>: t-2</td>
</tr>
<tr>
<td>: t-3</td>
</tr>
<tr>
<td>: t-4</td>
</tr>
<tr>
<td>Self-rated health excellent</td>
</tr>
<tr>
<td>Marriage status : divorced/sep</td>
</tr>
<tr>
<td>: never married</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Self-employed</td>
</tr>
<tr>
<td>Wave 8</td>
</tr>
<tr>
<td>Remaining standard controls</td>
</tr>
<tr>
<td>R^2 overall</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Individuals</td>
</tr>
</tbody>
</table>

Note: Waves 8&13.

**Summary of the new hypothesis tests on adaptation to income**

Although past income effects are found within the BHPS, particularly in relation to the first year lag, this finding is not especially robust, as noted from Table 6.5 where the first lag is no longer significant. The other hypotheses considered here are not well supported, and offer little help in explaining the negative finding on lagged income. Hence we cannot rule out adaptation to income, but this analysis suggests a more complex process than habitation to income might imply.

The classic idea of adaptation through expenditure patterns, tested here by considering the lagged income effect on those with and without mortgages, is not supported. Indeed, these findings point to a need for a greater understanding of the role of past income in wealth accumulation. The positive impact of past income upon financial coping and income satisfaction supports the notion of current period financial benefit from previous periods’ income. This suggests a need to look to other domains of life in order to understand the negative impact of past income upon life satisfaction. The accumulation of assets generates an opposing impact to
adaptation. In cases where the asset effect is strong, a negative impact of past income is less likely to be identified.

Signals and trends relating to the future clearly have a role to play in understanding these effects. The lagged income effect operates more strongly on those who experience declines in income, and feeling pessimistic about future finances is detrimental to life satisfaction. The latter may contain unobserved information relating to an individual’s financial status. Alternatively, it may point to the importance of personal interpretation of objective circumstances. Deterioration in income may also be related to self-esteem, self-confidence and personal judgements of one’s self worth.

Another possibility considered here is that higher past income generates attitudes and priorities that are not conducive to maximising SWB. However, this data does not lead to any clear results.

Analysis on an annual basis may not be suitable for understanding income adaptation effects. Any individual may not adapt to new levels of income in the short-medium term, yet each new generation may begin with a new set of income norms. If so, it is necessary to consider adaptation issues across a longer time span.

Further work is needed to understand the extent of adaptation to income and explore the possible future negative consequences of income rises. One avenue to explore is other non-observed variables that may relate to annual changes in income and life satisfaction. For example, decreases in income since the previous year may signal personal failures or unobserved problems, such as deterioration in the health of a member of the household.

6.3 Relative income effects, negative externalities and collectively overworking

Relative income effects have a long tradition within economics being hypothesised by Smith (1776), Marx (1891), Veblen (1899), Duesenberry (1949), Leibenstein (1950; 1962), Samuelson (1973: 218), Pollack (1976) and Frank (1984; 1985; 1997), among others. However, only more recently has empirical work given strength to the idea of relative income effects.

The previous chapter discussed empirical work exploring individual preferences towards alternative income distributions and the impact of relative incomes upon labour market decisions. Another strand of relative income research uses survey data to explore the link between subjective responses and relative income. This includes measures of income satisfaction, job satisfaction (Brown et al., 2007; Clark and Oswald, 1996; Groot and van den
Brink, 1999) and SWB. The focus here is on the use of SWB measures, therefore consideration will be restricted to the literature exploring the impact of others’ income upon SWB.

It has been argued (Layard, 1980; Ireland, 1998) that where incentives to work are based on a desire to gain relative status there will be negative externalities to individual income increases, such that a proportionate increase in income across the board leaves all individuals in the same position. This has been described as “perhaps the biggest negative externality in modern society” (Layard, 2005b: 148). If the non-status driven hours worked would achieve equilibrium, the incentive for status will result in an inefficient allocation of time towards work, and misallocation between work and other aspects of life (Ng, 1987). An implication of this zero sum game is that higher marginal tax rates improve outcomes. Layard notes that, “if income confers status, it can and should be taxed on efficiency grounds” (Layard, 1980: 738). The stronger the relative income effect, the more progressive an optimum taxation system (Boskin and Sheshinski, 1978). Work hours decisions under relative income effects can be presented as a classic prisoner’s dilemma.64 This is shown in detail in Appendix 6.A.

Where income inequality is high the average gap between an individual’s income and mean income will be larger than when inequality is low. The impact of others’ income may therefore be picked up through a negative impact of inequality within an individual’s reference group.

Duesenberry (1949) claims that people make upward comparisons and that having above average reference group income has no impact upon utility, reference income only impacting upon the individual when own income falls below reference income. Where there is asymmetry in the relative income effect this should strengthen the negative impact of inequality and further strengthen the case for progressive taxation to achieve allocative efficiency (Duesenberry, 1949: 103).

6.3.1 Current evidence on the impact of relative income

There is reasonable evidence that income has a positive although non-linear effect on SWB although the exact slope of the utility function at different incomes remains unclear (Dolan et al., 2006a; Clark et al., 2007). However, studies that find a significant effect of income or income rank within a country may be picking up an absolute income effect, or the impact of a recent change in income, or the importance of relative income. It is only recently that data has been available to provide fairly convincing evidence that income rank and social position matter over and above absolute levels of income. Reference effects may operate via income or

---

64 Frank (2004) illustrates an example of safety regulation under relative income competition using a prisoner’s dilemma framework.
consumption (McBride, 2001: 260). Whilst it is consumption that is revealed to others good quality consumption data is limited, hence most empirical work relies upon income.

Some evidence for a ranking effect can be gained by studies that control for real incomes and find an additional effect of a ranking variable. For example, Alesina et al., (2004) use the US GSS (1981-1996) and the Eurobarometer (1975-1992) to show that being in the top two income quintiles increases happiness and life satisfaction even when controlling for income. Knight et al., (2007) analyse a 2002 national survey in China and find that, controlling for household income, being in a household rated as having above average village income increases happiness, and being below average reduces happiness and that this relationship is monotonic.

Influential rank position may not be rank nationally, but rather rank at the state or local level, or within a group of people with whom the individual makes comparisons. It is possible to test the relative income hypothesis by considering the impact of changes to the individual’s rank within identified subgroups. Most studies that have explored relative income effects use a single nation, and estimate a reference income within a subgroup assumed to be the individual’s reference group.

Unfortunately, the identity of the reference group is not clear as people operate in a number of different spheres generating many possible, and indeed multiple, reference groups. The incomes and living standards of those within the family, neighbourhood, peer groups, work colleagues, perceived social class, region, country or worldwide could impact directly or indirectly upon well-being. The sociology literature tends to present social comparisons as those with whom one has spatial proximity and a high degree of interaction, and share similar opinions and abilities (Festinger, 1954). Comparison groups may be thought to comprise geographically placed concentric circles, the inner circle representing neighbourhood friends and contacts, the next circle representing work environment and accessible labour market, next a wider geographical area such as region, then nation wide, then international. However, the location of close friends and family may be on any of these circles.

Individuals are more likely to have knowledge of the income and consumption of those with whom they have regular contact. This would imply that the income of those closest to the individual, their partner, family and immediate peer group, would have the greatest impact. However, international travel and international media are likely to generate awareness of standards of living of those whom we do not come into direct contact, even those living in distant countries, which would suggest that the reference group - proximity relationship may have weakened since the 1950s.
Some studies have directly asked respondents with whom they make comparisons. Knight et al. (2007) report on a survey in China in which 68% of respondents gave their main comparator group as fellow-villagers, 7% as family, 4% the township and 7% China as a whole, with 14% not able to answer. Whilst most people may know with whom they make conscious income comparisons, given the myriad of ways in which others’ income can impact upon behaviour and evaluations, it is possible that people are unaware of the full impact of others’ income. For example, people may hold consumption desires without being aware that these have been influenced by the consumption of their friends or neighbours.

There are two main options for estimating an individual’s reference income. The first is to estimate predicted income from an income equation using individual characteristics (as done by Clark and Oswald, 1996). The second takes the average income from the sample of cases like the case in question. Group classifications used in the literature have included location, age, education level and gender. Average income data may be taken from within the same dataset or from a different dataset with similar information that is designed to be representative at the degree required for the reference group that can be linked to the individual. If the original dataset is used with narrowly defined subgroups there is a danger that there will be groups with few cases.

Reference income effects can be explored by considering the sign and significance of reference income in a typical SWB function (1), in which SWB of individual i at time period t is a function of own income (logYown) and reference income (logYref), a set of other social, economic and environmental factors (Z’), and an error term. If the coefficient on logYref (β2) is negative, this suggests that, controlling for own income, an increase in reference income lowers SWB.

\[
SWB_{it} = \beta_1 \log Yown_{it} + \beta_2 \log Yref_{it} + \beta_3 Z'_{it} + \epsilon_{it} \quad (1)
\]

The use of logs helps overcome the non-normal distribution of income, and where both own income and reference income are in logs it is easy to compare a percentage change in both. However, the logged structure for reference income assumes that the impact of an additional increase of 1% of reference income (holding own income constant) is the same at all levels of reference income (i.e. the absolute effect is stronger at lower levels of reference income).

Whilst the coefficient on reference income should only be significant if there is a relative income effect, the absolute and relative income effects are hard to determine. With reference income held constant a change in own income may reflect the impact of absolute income or relative income.
Findings from studies using this type of model (1) are indicative of a negative relative income effect. McBride (2001) explores the effect of relative income on overall happiness using data from the 1994 US GSS. When controlling for the log of own income he finds a negative yet insignificant effect for relative income, defined as the average income of ones age group from a 1994 population survey. Given that the dataset only gave 324 observations, and that the dependent variable was reported on a three point scale, this lack of significance is unsurprising, nevertheless, despite the large standard error, the size of the coefficient on reference income is considerably larger than that on own income. Additional analysis suggests the relative income effect is smaller at low-income levels where absolute income is more important. This would imply that relatives are a luxury good.

Clark (2003b) explores relative income effects using pooled data from the BHPS. He defines reference group income as the mean income of those of the same sex, region and year. Using a model equivalent to (1) he finds reference income to have a significantly negative influence on life satisfaction for a group of full-time employed workers when controlling for own income. Reference income is not, however, significantly related to the GHQ. In both cases the size of the coefficient on reference income is large compared to that on own income, although it is generally poorly defined.

Stronger evidence in support of a negative relative income effect comes from Luttmer (2005) who finds that the log of predicted average earnings in the individual’s local area (defined from US Public Use Microdata Areas (PUMAs) which contain on average 150,000 inhabitants) is negatively related to overall happiness (although not to depression). He uses happiness data from the 1987-88 and 1992-94 waves of the US National Survey of Families and Households (NSFH). Average annual earnings were estimated for the PUMAs by applying national earnings by industry, occupation and year from the Current Population Survey to the industry and occupation mix of that PUMA.

These findings are robust to a range of specifications, including controlling for state, time and individual fixed effects and using predicted rather than actual income based only on those individual controls that do not vary with local prices. Luttmer also considers school quality and changes to the housing market that could both impact on happiness. However, different effects are not found by homeownership, or the presence of children in the household. Those who socialise with their neighbours less than once a month, and hence would be less exposed to neighbourhood comparison incomes, are less affected by reference income, yet differences in socialising outside of the neighbourhood has no impact, reinforcing that this effect is due to social comparison rather than potential differences in living costs between the areas. The coefficient on household income is only slightly bigger than that on relative earnings suggesting
“an increase in neighbours’ earnings and a similarly sized decrease in own income each lead to a reduction in happiness of about the same order of magnitude” (Luttmer, 2005).

Weinzierl (2005) uses West German GSOEP data and three year moving averages of equivalent log household income to find a negative sign for the log of reference income. In this case the reference group is determined by respondents of a similar age (five-year birth groups, limited to those between 26 and 60), gender and education level. He also finds that the positive coefficient on log of income is virtually matched by the negative coefficient on log of peer group income.

The relative income effect becomes more explicit when it is modelled using the difference between an individual’s income and the average reference income \( Y_{\text{difference}} \) as in (2).

\[
SWB_{it} = \beta_1 Y_{\text{difference}} + \beta_2 \log Y_{\text{ref}} + \beta Z_i + \alpha_i + \epsilon_i
\]  

(2)

Including the difference between own and reference income, when controlling for reference income, provides the same information as including own and reference income separately.\(^{65}\)

Holding constant reference income, if the difference between own income and reference income increases the individual has moved up the ranking. If own income and reference income increase by the same amount, the difference will remain the same, but reference income will increase. If absolute incomes matter this should give a positive sign to the reference income coefficient. If only relative income matters, the coefficient on reference income should be zero. If only absolute income matters, the coefficients on own and difference income should be both positive and identical.

Own income rather than reference income could be included, but this is likely to be problematic if the difference is highly correlated to own income (which is likely to be the case).

Studies using a difference model, also find a negative relative income effect. Dorn et al., (2007) find that for Swiss and international data, controlling for individual fixed effects, reference income (defined as 40% of average income in the Canton for Switzerland, or national average equivalent income for the international data, not logged) does not have a significant effect on overall happiness but the difference between an individual’s own income and reference income is positively related to overall happiness. This supports a relative income hypothesis, however,

\(^{65}\) The difference income is the log of own income less the log of reference income, so (2) can be rearranged to the structure of (1) as shown in (3) and (4).

\[
SWB_{it} = \beta_1 (\log Y_{\text{own}} - \log Y_{\text{ref},t}) + \beta_2 \log Y_{\text{ref}} + \beta Z_{it} + \epsilon_i
\]  

(3)

\[
SWB_{it} = \beta_4 \log Y_{\text{own},t} + (\beta_2 - \beta_4) \log Y_{\text{ref}} + \beta Z_{it} + \epsilon_i
\]  

(4)

If the coefficient on the difference between own income and reference income \( (\beta_4) \) is the same as the coefficient on average reference income \( (\beta_2) \), this effectively removes reference income (as can be seen from (3) and (4)).
in neither case can they reject the possibility that the coefficient on reference income ($\beta_2$) and the coefficient on the difference between own and reference income ($\beta_4$) are the same, hence they are not able to rule out an absolute income effect. Dorn et al., (2007) also include a term for the square of the income difference where it is positive, which they find to have a significantly negative effect in both datasets, suggesting diminishing returns to relative income.

Ferrer-i-Carbonell (2005) presents the German data (1992-1997) in both a log-linear and difference structure. In the log-linear model, the coefficient on the log of reference income (based on cell averages for region, age and education, and for some models, gender, but this inclusion does not change the result) is negative and very similar to that on household income. Using the difference structure, she finds a positive effect for the difference between log of own income and log of reference income. This is significant for the whole sample, West Germany (at 10%) but not East Germany, which may be due to increased instability in Eastern Europe resulting in relative income giving more mixed messages. She concludes that, “increases in family income accompanied by identical increases in the income of the reference group do not lead to significant changes in well-being” (Ferrer-i-Carbonell, 2005: 1015). She also finds that the impact of reference income is asymmetric. Being above one’s reference income has a positive but non-significant effect on overall happiness, being below one’s reference income has a negative and significant effect, suggesting that upward comparisons matter.

Graham and Felton (2006) also present Latin American data in a log-linear and difference model. Using a difference model they find that controlling for average wealth the difference between individual wealth (an index of ownership of 11 goods) and average wealth (at the country or city size level) is positive and significant and that the coefficient on average wealth is not significant, again suggesting a relative wealth effect.

An alternate structure is to model the ratio of household income to average reference income:

$$ SWB_u = \beta_5 \log Y_{own, u} + \beta_6 \frac{\log Y_{own, u}}{\log Y_{reference, u}} + \beta_7 Z_{u} + \epsilon_u$$  \hspace{1cm} (5) 

Again it may be hard to get well-determined coefficients on this model if the log of own income is included as well as the ratio rather than reference income if own income is highly correlated with the own/reference income ratio.

66 She uses an ordered probit model with individual random effects controlling for the individual mean of log income, log education, log of the number of children and log of the number of adults (i.e. a Mundlak approach).
Hudson (2006) adopts this structure using European data (Eurobarometer 2001, without a control for reference income) and found that the ratio of household income to average country income has a positive effect on life satisfaction.

When Blanchflower and Oswald (2004) model the US GSS (1972-1998) data as a ratio, unlike the non-significant findings when using model (1), the ratio of individual income to state income per capita has a significant and positive effect on overall happiness even when controlling for household income, household income per capita and regional house price index (to control for state level costs of living). They also find that the ratio of own income to the 5th income quintile has the largest effect, suggesting that people make upward comparisons. When also controlling for the ratio of individual income to the 1st and 2nd quintile average these have a negative effect hence being further away from the poorest end of the distribution is detrimental, suggesting a negative role for income inequality. Also using the US GSS (1989-1996) Hagerty (2000) finds that maximum area income and the 80th percentile income in the area reduces happiness. Ball and Chernova (2007) find that controlling for logged own income the log of the ratio of own to median income in 18 countries from the WVS (1995-1998) is positive and significant, and that a percentage change in relative income position is two to three times more influential on SWB than a percentage change in absolute income.

Senik (2004a) finds quite contrasting results from the RLMS (1994-2000). She uses format (1) and predicted income for reference income, based on productive characteristics (using the same dataset). The log of reference income is positively related to life satisfaction when controlling for household and individual income (using a range of specifications and controlling for fixed effects). Senik links this to the potential for higher relative incomes to transmit information about an individual’s future, since knowing that others similar to oneself are currently earning more may lead people to be more optimistic about their own future earnings. Given the uncertainty of the Russian economy, relative incomes may confer particularly important messages. Hirschman and Rothschild (1973) use the phrase ‘tunnel effect’\(^\text{67}\) to describe situations where individuals may “draw gratification from advances of others”, during high growth stages of an economy, at least “for a while” (Hirschman and Rothschild, 1973: 546).

However, Senik’s finding may also arise if own income has high measurement error, which may be more likely in a transition economy. If the relationship between own income and SWB is being obscured by noise in the income measure, then the coefficient on own income will be bias towards zero. However, the measurement error on relative income is likely to be lower because individual inaccuracies will average out. If reference income is positively correlated with the

\(^{67}\) Stationary drivers waiting to enter a tunnel may interpret the progression of an adjacent line of traffic as a positive sign that the blockage is being cleared.
individual’s true level of income, this will cause the regression to attribute some of the positive relationship between own income and SWB to the reference group income (Weinzierl, 2005). Senik finds that the positive influence of reference income is stronger the more uncertain agents are about their professional and material future and the higher their income volatility. It is also stronger for younger individuals (<40) and those whose professional future is longer. Whilst income volatility may point to measurement error, the stronger effect on the young and those with uncertain futures also gives weight to a ‘tunnel effect’ explanation.

Senik (2004b) also finds a positive influence of reference income on satisfaction with income and satisfaction with future prospects in the Hungarian TARKI panel (1992-1997). The result is stronger for those under 40, and for risk adverse people. A similar result is found in the Polish Household Survey (1994-2000) and household surveys in Estonia, Latvia and Lithuania (Senik, 2004b).

Kingdon and Knight (2007) also find interesting results in their analysis of data from South Africa. They analyse responses to a question that asks how satisfied the household is with the way it lives these days. The log of average household income per capita of the community or cluster (each cluster contains about 580 households) is positively related to satisfaction. Whereas average income at the district level has a negative, but non-significant coefficient, unless race-based averages are considered. The authors suggest four possible explanations. The first, risk sharing, is considered unlikely because low risk white households share the same relationship to cluster income. The second, a ‘social wage’ with the higher cluster income picking up the quality of publicly provided goods such as education and sanitation, is also considered unlikely because social amenities did not explain the coefficient on cluster dummies (although this did not include security variables). A third possibility relates to measurement error in the income variable, as discussed above. This is explored by regressing household income on a set of explanatory variables, and then adding average cluster income. Whilst this addition improves the model, the difference is not such that it suggests a high level of measurement error in household income (R squared increases from 0.558 to 0.656). The fourth possibility, the presence of altruism, is seen as offering the most promising explanation. When clusters are divided into those with under 200 households, the authors find that for these households average cluster income has a positive effect and that cluster income is insignificant for larger clusters.

Knies et al. (2007) also find evidence of a positive (but non-significant) relative income effect within the German GSOEP data (1994&1999) drawing neighbourhood income data from matched zip code (average size of about 10,000 individuals). A range of covariates are included in this study including house and community type and access to public facilities within the
neighbourhood. The authors suggest the positive finding may arise from higher income areas signalling improvement and possibly attracting new business, residents of these areas avoiding stigmatisation based on home location particularly in the labour market, and high income areas having a more pleasant and safer environment if wealthier individuals invest more in maintaining their properties.

Another approach to exploring relative income effects is to consider the impact of income inequality, since the greater the inequality, the more intense will be relative deprivation. Furthermore, as noted above, if relative income effects are asymmetric this would also contribute to a negative impact of inequality.

Cross-country evidence on income inequality and SWB has mixed findings. For example, Fahey and Smyth (2004) find that inequality reduces life satisfaction in their analysis of the WVS (33 societies, 1999-2000). Hagerty (2000) also finds a negative relationship with income skew, the Gini coefficient and the percentage of national income received by households in the highest income quintile, on the average of country life satisfaction, from analysis of 8 countries (various years 1972-1994). On the other hand Haller and Hadler (2006), use data from 41 countries from the WVS 1995-1997 and find that income inequality (Gini coefficient) increases life satisfaction.

One explanation for these contrasting findings using international data may be that the inclusion of particular countries influences the results (Bjørnskov, 2003). Specifically, the relatively happy Latin American countries tend to have fairly unequal income distributions, and relatively unhappy former-Communist countries tend to have fairly equal income distributions. Haller and Hadler (2006) suggest high happiness in Latin American countries may result from high quality social relations offsetting adverse macro-social conditions and that economic development in Latin American countries was fairly dynamic during the years they analysed. Nevertheless within Latin American countries income inequality and happiness are found to be strongly negatively correlated (Graham and Felton, 2006).

Alesina et al. (2004) find that the impact of inequality varies between Europe (Eurobarometer, 1975-1992) and the US. Inequality in the US, measured by the Gini coefficient, only significantly reduces overall happiness for the rich and politically left-leaning. In Europe, however, inequality reduces life satisfaction, particularly for those with left wing political leanings and the poor.

Using the same Eurobarometer data but fewer controls and fewer years (1995-1998) O’Connel (2004) finds that income inequality is positively related to life satisfaction. However, the
Alesina et al. (2004) findings are supported by Schwarze and Harpfer (2007), who find that inequality (Gini coefficient) within West German regions (GSOEP, 1985-1998) reduces life satisfaction, including when individual fixed effects are controlled for. When the sample is split by income tercile, however, the effect is only significant for those in the lowest income tercile.

For Britain, Clark (2003b) finds that for full-time employed individuals, income inequality (Gini coefficient) in one’s reference group (based on gender, region and year) increases life satisfaction, particularly for those under 40, those on below average incomes and those who have experienced a greater increase in income over the last three years. However, the impact on the GHQ is mostly insignificant, and controlling for fixed effects inequality only increases life satisfaction for those under 40. Senik (2004) finds that for Russia (RLMS, 1994-2000) regional income inequality (Gini and Stark index) does not have a significant effect on life satisfaction. Knight et al., (2007) find that for China inequality (Gini coefficient) in the county (about half a million people) is significantly positively related to happiness. Biancotti and D’Alessio (2007) find the impact of income inequality to vary according to individual values of moderation and inclusiveness in ESS 2004 data.

These findings suggest that the effect of income inequality varies depending on the how inequality is interpreted. Clark (2003b), Alesina et al. (2004) and Knight et al. (2007) explain their findings in terms of inequality communicating messages of opportunity. The message from income inequality is likely to vary according to perceptions of social mobility (Benabou and Triole, 2006). Where mobility is perceived to be low, such as Europe and Germany, inequality is found to have a negative impact.

Inequality may also have an indirect impact upon other determinants of SWB. For example, income inequality has been linked to higher working hours (Bowles and Park, 2005), the erosion of social capital (Kawachi et al., 1997) and poor health (Wilkinson, 1996). However, regional inequality in the UK has not been found to be significantly related to health (Lorgelly and Lindley, 2007), or psychological health (Wildman and Jones, 2002). In summary, there is reasonable evidence that relative income effects are important, and that the nature of this comparison is upward looking. However, if the relative income effect is robust one would expect it to be detected using different representations of the data, and identified in a range of different datasets. Whilst studies using different data and slightly different models mostly show that relative income impacts upon behaviour, domain satisfaction and SWB, neither the significance, nor the direction, of these finding are consistent across studies.

Part of the variation across these studies may arise due to differences in the way the variables are modelled. Some studies use individual income and individual reference income, others use
household; some use equivalised income others do not; some use logged income others do not; some base reference income on predicted income for an individual with those characteristics others use cell averages, which themselves are based on a variety of characteristics including location (the size of area varying considerably), age, education, gender; some use a log linear model, others a difference or ratio, for the latter some control with own income others with reference income; the other independent variables included differ between studies, variables which may be related to the income variables (such as income inequality, lagged income, income growth, education and health) are variably included, similarly, not all studies control for individual effects, or geographical location. Furthermore, studies use different dependent variables to measure SWB which may pick up different attributes of SWB and vary in terms of response options.

Another possible explanation for the apparent lack of consistency may be that the composition of samples differs between studies and the impact of income and others’ income varies depending on an individual’s personal circumstances and personal interpretation of the absolute and relative income.

There is some evidence that the strength of the relationship between income and SWB depends upon an individual’s circumstances. Income effects are thought to operate more strongly at lower levels of income (Fahey and Smyth, 2004; Helliwell, 2003). McBride (2001) suggests reference income matters less at lower incomes where absolute income is more important. Income may be more important at certain life stages (Cummins et al., 2004; Gerlach and Stephan, 1996; Pichler, 2006), although Marks and Fleming (1999) find no significant interaction between age and income.

There is also some evidence that reference income matters more to certain groups. Luttmer (2005) found that the effect was strongest for those aged 30 to 60 and for married or divorced respondents, but similar between men and women, those with and without children, and those renting or homeowners. For the US, Blanchflower and Oswald (2004) do not find any significant difference between the effect on men and women, however, they find for the UK that being in the 2nd and 3rd income quartiles is more beneficial to women. In Germany Di Tella et al. (2007) find that household income matters more for men, and left-wingers.

Whist the evidence is suggestive of a relative income effect, in some cases unobservable, possibly geographic components may be influencing findings. Support against the relative income effect picking up price differences can be drawn from Luttmer’s (2005) finding that both renters and home owners experience a similar relative income effect, and that the effect is stronger for people who socialize more with their neighbours, but not for those who socialize
more with friends outside the neighbourhood. Blanchflower and Oswald (2004) also find that their results are robust to the inclusion of regional house price index.

Another possibility is that relative income effects arise due to selection effects if selecting into a particular geographic area is influenced by SWB. However, this could operate in either direction. Unhappy people may be more status orientated and seek the postcode status of a high-income area, alternatively, the majority of movers may seek the higher income areas, but the happier people may be more motivated to actually move.

If the use of the SWB scale is not independent of income (Osmani, 1993) this would undermine efforts to identify the true relationship between income and SWB. People may respond to questions about happiness relative to a happiness norm that may be income dependent, hence what is meant by 8/10 on a life satisfaction scale may be different for a rich person compared with a poor person. Whilst this may limit confidence in the actual income coefficient, it should not undermine findings relating to the impact of others’ income or income inequality. However, Luttmer raises a concern that living in a prosperous area might affect one’s definition of happiness even if it does not affect one’s true or experienced well-being. A happy person living in a rich area may in fact be happier than a happy person in a poor area. Luttmer addresses this by looking at the frequency of financial worries and marital disagreements about money which he argues to be more objective scales less subject to scale re-norming based on area income.

In 1997 Frank gave the following summary relating to reference income effects,

“In sum, the claim that satisfaction depends heavily on relative position is supported by considerable evidence from both the psychology literature on subjective well-being and by at least fragmentary evidence from the behavioural economics literature. I am aware of no empirical or theoretical evidence against this claim” (Frank, 1997: 1836).

Since that date, evidence showing a relative income effect has accumulated, yet so too has evidence which suggests that in some circumstances there may be positive relative income effects. The relationship between income and SWB is extremely complex, and closely tied up with an individual’s perceptions and expectations, which are themselves likely to be dynamic and respond to the environment in which the individual operates. There is much still much to understand before a clear conclusion can be drawn on the magnitude of reference income and the reference points people use when making income comparisons.

**6.3.2 New hypothesis on the impact of relative income**

Negative externalities from increases in others’ income can potentially affect all people, however, a change in marginal income tax would impact upon those currently working. If working individuals gain no personal benefit from improvements in relative income position,
their additional income may still impose externalities upon others. Externalities generated by relative income effects that they themselves do not experience will not lead to an inefficient work/leisure balance for the individual. It is therefore necessary to clearly identify the magnitude of relative income effects upon those who are employed and are at risk of an inefficient allocation of time between work and leisure if relative income effects are to be successfully incorporated into taxation policy. Unfortunately, existing work does not present a clear, robust effect of relative income upon well-being. Three weaknesses in existing work will be explored here.

Firstly, although some consideration has been given to costs differences according to location this has been fairly limited. Some studies have controlled for costs using regional house prices, however, the regional house price index may not be the most ideal means of controlling for regional price differences, since this has both cost and wealth consequences for households. Specific housing costs in addition to regional house prices are used here. The correlation between income and housing costs is likely to lower the absolute income effect but should help isolate the relative income effect from local cost factors.

*Hypothesis i: Relative income effects are robust to the inclusion of household housing costs against a null of non-significance once housing costs are included.*

Secondly, many models do not control for past income. Living in a higher income area may suggest an increased likelihood of high past income. Given the negative effect of past income discussed in 6.2 it is important to ensure that this is controlled for.

*Hypothesis ii: Relative income effects are robust to the inclusion of past income against a null of non-significance once past income is controlled for.*

Thirdly, it is not clear whether a change in relative income has the same impact whether it arises from a change in own income, or a change in others’ income. The latter may arise either by a change in average income of one’s current area (which is likely to be a gradual effect) or through moving to a new geographical area (which is likely to be sudden). Distinguishing between those who move region and those who do not will help to isolate the types of individuals and households who experience a relative income effect.

*Hypothesis iii: Controlling for own income the impact of regional average income will differ when the individual has changed regions compared to not changing regions, against a null of no difference.*

As noted above a strong relative income effect would lead us to expect a negative coefficient on income inequality. The impact of inequality using the BHPS has been explored by Clark (2003b) who finds a positive effect on regional wage income inequality (of one’s own gender).
for full time employees, significantly for the employed under 40 year olds once fixed effects controlled for. However, the BHPS can also be used to consider the impact of inequality more generally using a wider sample, and using income inequality rather than wage inequality. This is important, partly because of conflicting evidence in relation to inequality and partly because of the political relevance of this knowledge.

**Hypothesis iv:** Income inequality has a negative impact upon SWB, against a null of non-significance or positive impact.

However, as indicated by other work the impact of inequality varies by values such as moderation and inclusiveness (Biancotti and D’Alessio, 2007), by age and gender (Clark, 2003b), by political persuasion (Alesina et al., 2004), and by income level (Alesina et al., 2004; Schwarze and Harpfer, 2003). It is important to gain a better understanding of these subgroup effects. If these different effects are caused by whether the individual stands to benefit from the surrounding inequality, we would expect noticeable differences in the impact of income inequality across the life course. Young people would be more likely to see themselves as beneficiaries, whereas the retired and old, without any prospect of catching up, would be more likely to see themselves as losers.

**Hypothesis v:** The impact of income inequality varies across the life course against a null of no difference.

### 6.3.3 Testing new hypothesis on the BHPS

BHPS data confirm the finding of a relative income effect. Table 28 shows coefficients from an OLS fixed effects model, using standard controls (similar to 2.2 above\(^{68}\)) and both a log-linear model (1) and ratio model (5) set out in 6.3.1. Regional average income (net current income averaged for the 18 regions by wave, excluding the individual) has a strongly negative impact upon life satisfaction, with a surprisingly large coefficient compared with the coefficient size for own income. Indeed once the relative income ratio is controlled for own income has a significantly negative impact upon life satisfaction, suggesting that relative position is more important than absolute income.

The high correlation between own income and the ratio of own to mean income (0.972) may be creating problems of multicollinearity. Consequently, the model is re-run using mean regional income and the relative income ratio (the fifth and sixth column). In this model the size of the

---

\(^{68}\) Regional dummies are not included since the aim is explore a regional effect, however, whether the region is a metropolitan area is controlled for. The inclusion of own income and the ratio of own to regional income would give a constant ratio between these two variables within each region, thereby generating a collinearity problem (see Ball and Chernova, 2007).
The ratio coefficient is much reduced. Mean regional income remains similar in size and negative. These findings suggest that a proportionate increase in all incomes does not enhance life satisfaction.

The influence of regional cost differences are explored through the introduction of a regional house price index (data from The Nationwide based on regions, but not disaggregated into metropolitan and non-metropolitan areas) and household housing expenditure (2nd, 4th and 6th columns of Table 28). Both have the expected sign but are insignificant and do not substantially alter the impact of relative income. Consequently, the hypothesis that relative income effects are not robust to the inclusion of housing costs can be rejected.

Table 28: OLS-FE on life satisfaction, including relative income

<table>
<thead>
<tr>
<th></th>
<th>Additive model</th>
<th>Ratio model I</th>
<th>Ratio model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log own hh income</td>
<td>.0570</td>
<td>-.28372</td>
<td>-.2956</td>
</tr>
<tr>
<td></td>
<td>(.0172)**</td>
<td>(.1284)*</td>
<td>(.1288)*</td>
</tr>
<tr>
<td>Log mean regional hh income</td>
<td>-.3267</td>
<td>-</td>
<td>-.2689</td>
</tr>
<tr>
<td></td>
<td>(.1318)*</td>
<td></td>
<td>(.1320)*</td>
</tr>
<tr>
<td>Ratio own hh income/mean regional hh income</td>
<td>.0189</td>
<td>.0194</td>
<td>.0032</td>
</tr>
<tr>
<td></td>
<td>(.0071)**</td>
<td>(.0071)**</td>
<td>(.0010)**</td>
</tr>
<tr>
<td>Regional house prices</td>
<td>-1.53e-06</td>
<td>-1.50e-06</td>
<td>-1.50e-06</td>
</tr>
<tr>
<td></td>
<td>(9.33e-07)</td>
<td>(9.33e-07)</td>
<td>(9.33e-07)</td>
</tr>
<tr>
<td>Net housing costs</td>
<td>-0.000046</td>
<td>-0.000046</td>
<td>-0.000046</td>
</tr>
<tr>
<td></td>
<td>(.00005)</td>
<td>(.00005)</td>
<td>(.00005)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0280</td>
<td>0.0280</td>
<td>0.0280</td>
</tr>
<tr>
<td>Observations</td>
<td>39,675</td>
<td>39,675</td>
<td>39,675</td>
</tr>
<tr>
<td>Individuals</td>
<td>10,616</td>
<td>10,574</td>
<td>10,574</td>
</tr>
</tbody>
</table>

Note: Employed or self-employed only. Waves 7-10&12-14. Coefficient(se).

Table 29 shows the results when past income is included. Although reference income is no longer significant within the additive or ratio model (1) this is due to the sample, which is reduced due to availability of cases with three years of lagged income.

Table 29: OLS-FE on life satisfaction, including past and relative income

<table>
<thead>
<tr>
<th></th>
<th>Additive model</th>
<th>Ratio model I</th>
<th>Ratio model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log own hh income</td>
<td>.0660</td>
<td>-.1167</td>
<td>-.1140</td>
</tr>
<tr>
<td></td>
<td>(.0212)**</td>
<td>(.1489)</td>
<td>(.0102)</td>
</tr>
<tr>
<td>Log mean regional income</td>
<td>-.2218</td>
<td>-.2249</td>
<td>-.1555</td>
</tr>
<tr>
<td></td>
<td>(.1534)</td>
<td>(.1536)</td>
<td>(.1534)</td>
</tr>
<tr>
<td>Ratio own hh income/mean regional hh income</td>
<td>0.1014</td>
<td>.0102</td>
<td>.0036</td>
</tr>
<tr>
<td></td>
<td>(.0083)</td>
<td>(.0083)</td>
<td>(.0012)**</td>
</tr>
<tr>
<td>Log own hh income : t-1</td>
<td>-.0228</td>
<td>-.0227</td>
<td>-.0227</td>
</tr>
<tr>
<td></td>
<td>(.0194)</td>
<td>(.0194)</td>
<td>(.0194)</td>
</tr>
<tr>
<td>Log own hh income : t-2</td>
<td>-.0299</td>
<td>-.0298</td>
<td>-.0299</td>
</tr>
</tbody>
</table>
Chapter 6  Income and SWB: policy implications

Table 30: OLS-FE on life satisfaction, including relative income, for ‘movers’ and ‘non-movers’

<table>
<thead>
<tr>
<th></th>
<th>Additive Model</th>
<th>Ratio Model I</th>
<th>Ratio Model II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Movers Non-movers</td>
<td>Movers Non-movers</td>
<td>Movers Non-movers</td>
</tr>
<tr>
<td>Log own hh income</td>
<td>.0445 (.0452)</td>
<td>-.8856 (.2307)**</td>
<td>.0802 (.1539)</td>
</tr>
<tr>
<td></td>
<td>.0591 (.0186)**</td>
<td>(.1584)</td>
<td></td>
</tr>
<tr>
<td>Log mean regional income</td>
<td>-.9627 (.2366)**</td>
<td>.0645 (.1584)</td>
<td>-9178 (.2367)**</td>
</tr>
<tr>
<td></td>
<td>(.196)</td>
<td>(.1584)</td>
<td></td>
</tr>
<tr>
<td>Ratio own hh / regional mean income</td>
<td>.0519 (.0129)**</td>
<td>-0.0012 (.0085)</td>
<td>.0025 (.0025)</td>
</tr>
<tr>
<td></td>
<td>(.0085)</td>
<td>(.1584)</td>
<td></td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes Yes Yes Yes Yes Yes</td>
<td>Yes Yes Yes Yes Yes Yes</td>
<td></td>
</tr>
<tr>
<td>R^2 within</td>
<td>0.050 0.0254 0.0499 0.0274 0.0500 0.0274</td>
<td>0.0500 0.0274</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>5402 34,273 5402 34,273 5402 34,273</td>
<td>5402 34,273</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>1746 8870 1746 8870 1746 8870</td>
<td>1746 8870</td>
<td></td>
</tr>
</tbody>
</table>

Note: Employed or self-employed, waves 7-10&12-14. Coefficient(se).

Table 30 clearly identifies the relative income effect to be confined to those who move regions. A gradual change in relative income does not appear important. The strong impact upon those who move regions suggests migrants might be particularly susceptible to reduced life satisfaction arising from relative income effects.

Whilst the BHPS data suggest that relative income is important, at least for some groups, the size of the relative income effect remains unclear. If a clear indication of the size of the relative income effect could be gained there is still little evidence to show how relative income influences the work life balance and the extent to which it results in collective overworking.

---

69 The difference between the coefficients is supported by a Wald test.
A strong relative income effect should result in a negative impact of income inequality within the individual’s reference group. Inequality is measured here at the regional level, using the Gini coefficient, the Theil index, and the coefficient of variation (Cowell, 2000). Controlling for lagged income and regional average household income, each inequality measure shows a negative sign, although none are significant. Consequently, we cannot reject the null of the fourth hypothesis that controlling for own income, and regional average income, regional inequality has no significant effect on life satisfaction.

Table 31: OLS-FE on life satisfaction, including income inequality

<table>
<thead>
<tr>
<th></th>
<th>Gini</th>
<th>Theil</th>
<th>Coefficient of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log own hh income</td>
<td>.0411(.0149)**</td>
<td>.0413(.0150)**</td>
<td>.0413(.0150)**</td>
</tr>
<tr>
<td>Log mean regional hh income</td>
<td>-.2056(.1305)</td>
<td>-.2047(.1380)</td>
<td>-.2023(.1349)</td>
</tr>
<tr>
<td>Regional hh income inequality</td>
<td>-.3607(.2887)</td>
<td>-.1427(.1438)</td>
<td>-.0459(.0404)</td>
</tr>
<tr>
<td>Lagged own hh income</td>
<td>-.0129(.0132)</td>
<td>-.0129(.0132)</td>
<td>-.0129(.0132)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0398</td>
<td>0.0398</td>
<td>0.0398</td>
</tr>
<tr>
<td>Observations</td>
<td>54,666</td>
<td>54,666</td>
<td>54,666</td>
</tr>
<tr>
<td>Individuals</td>
<td>13,161</td>
<td>13,161</td>
<td>13,161</td>
</tr>
</tbody>
</table>

Note: Waves 7-10&12-14. Coefficient(se).

The average effect of inequality masks some interesting differences. From a policy perspective the average effect will be paramount, however, in terms of understanding the processes involved subgroup analysis is essential.

In the BHPS the impact of inequality varies across the life cycle (Table 32). At a young age (<25) inequality is insignificant but positive. At a young-middle age, during the time that many people purchase property and bring up children, inequality has a significant negative impact. At old-middle age (from 50 to 70) the effect becomes positive although reference income is significantly negatively. The effect of inequality changes to negative again beyond 70.

Table 32: OLS-FE on life satisfaction, by age group

<table>
<thead>
<tr>
<th></th>
<th>Age &lt; 25</th>
<th>Age 26-50</th>
<th>Age 51-70</th>
<th>Age &gt;70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log own hh income</td>
<td>.0318(.0808)</td>
<td>.0594(.0230)**</td>
<td>.0155(.0239)</td>
<td>-.0316(.0536)</td>
</tr>
<tr>
<td>Log mean regional hh income</td>
<td>-.2179(.7011)</td>
<td>-.00747(.1813)</td>
<td>-.6994(.2455)**</td>
<td>-.0515(.4034)</td>
</tr>
<tr>
<td>Regional hh income inequality (Gini)</td>
<td>1.265(1.5302)</td>
<td>-.8848(.3936)*</td>
<td>.5714 (.5357)</td>
<td>-1.627(.9319)</td>
</tr>
<tr>
<td>Lagged own hh income</td>
<td>.0128(.0590)</td>
<td>-.0203(.0205)</td>
<td>.0007(.0218)</td>
<td>-.0539(.0476)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R² within</td>
<td>0.0578</td>
<td>0.0445</td>
<td>0.0327</td>
<td>0.0554</td>
</tr>
<tr>
<td>Observations</td>
<td>4,060</td>
<td>27,116</td>
<td>15,621</td>
<td>7,869</td>
</tr>
</tbody>
</table>

70 The lack of significance may arise if regional inequality is not the level at which inequality impacts upon well-being, as suggested by Wilkinson and Picket (2006).
It should be noted that even if inequality does not show a negative impact when income is controlled for, under diminishing returns to income high income inequality would suggest a failure to maximise aggregate well-being. Most empirical models of SWB assume diminishing marginal returns to income, captured by allowing a concave form, either by logging income (hence assuming an elasticity of one), or including a polynomial. Layard et al. (2007) test the elasticity of the marginal utility of income using a range of different datasets and conclude that an elasticity of one is a reasonable approximation, and if anything slightly understates the extent of diminishing returns to income.

For any given level of regional average income, greater regional inequality will result in a greater number of individuals with below average regional income. Where there are diminishing marginal returns to income we would expect this to result in a lower SWB than a less unequal, mean preserving distribution. Using the BHPS data we can see that when own income is not controlled for higher income inequality is detrimental to individual SWB, either controlling or not controlling for average regional income, although this is only significant for the latter.

Table 33: OLS-FE on life satisfaction, income relative income and income inequality, but not controlling for income

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean regional income</td>
<td>-.0007(.0004)</td>
<td>-.4547(.2647)</td>
</tr>
<tr>
<td>Regional income inequality (Gini)</td>
<td>-.4966(.2261)*</td>
<td>-.4547(.2647)</td>
</tr>
<tr>
<td>Other controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R^2 within</td>
<td>0.0399</td>
<td>0.0396</td>
</tr>
<tr>
<td>Observations</td>
<td>71,718</td>
<td>64,056*</td>
</tr>
<tr>
<td>Individuals</td>
<td>16,392</td>
<td>15,258</td>
</tr>
</tbody>
</table>

Note: Waves 7-10&12-14. The second equation has fewer observations because the estimation for regional income excludes the individuals household income therefore is missing for those without household income data. Coefficient(se).

6.4 Individuals may over-consume due to relative consumption effects

In addition to important possible implications upon working hours, relative income and consumption effects may also generate inefficiencies if they lead to excess consumption and insufficient savings. Duesenberry (1949) argues that because people are mostly upward looking when making social comparisons, their perceived needs and aspirations are typically above their available resources. He presents wants as dynamic and determined by external influences,
including the consumption of others, which create a “demonstration effect”. People become dissatisfied with inferior goods once they have witnessed superior goods.

The consumption and income of others may lead to feelings of dissatisfaction, as identified in 6.3. It may also lead to non-optimal behaviour. A desire to ‘keep up’ or ‘catch up with the Joneses’ may result in excess consumption in the current period at the expense of future consumption. Pressure imposed by the consumption of others may lead to individuals failing to maximise over the long term due to weakening of their self-control or encouraging them to overlook the fact that relative consumption will still impact in the future.

Frank (1999) has argued that over consumption and low savings rates in high income countries arises due to reference group effects, particularly in relation to children’s education, suggesting the need for a progressive consumption tax to reduce the negative impact of reference group consumption upon desires and aspirations.

### 6.4.1 Current evidence on relative income effects and over-consumption

There is some empirical work to suggest that consumption is influenced by the consumption of others. For example, Duesenberry (1949) finds that the percentage of income spent on consumption is highly correlated with the person’s rank order in the local income distribution group. In support of Duesenberry’s ‘demonstration effect’, reference groups are found to influence the products and brands that people purchase (Bearden and Etzel, 1982; Childers and Rao, 1992). Using a 1980/81 survey from the Netherlands, Alessie and Kapteyn (1991) find that reference group consumption (defined by those of the same age, education and occupation) has an impact upon consumption. Medical care is the most conspicuous (however the authors note that this may arise due to differences in employee medical insurance), followed by education and recreation, clothes and footwear, transport, then housing and food.

If higher reference group consumption results in pressure to consume this would be expected to lead to lower savings and potentially increased debt amongst those exposed to higher reference income.

Kapteyn (2000) uses data from the Netherlands (1994-1998) to explore this issue. He finds that ‘having friends or acquaintances with high income tends to reduce savings” (Kapteyn, 2000: 11). However, he notes that this analysis may suffer from endogeneity if those who spend more associate with people with high incomes, or are more likely to report having acquaintances with high incomes or high spending.
Morgan and Christen (2003) analyse US data (1980-2000) and find a strong positive effect of income inequality on household debt. Consumer borrowing is found to be more sensitive to changes in income inequality than to changes in interest rates. The level of non-revolving debt (e.g. credit card), which is primarily used to finance consumer durable purchases, is most affected by income inequality.

There is therefore some evidence that relative income and consumption effects result in increased consumption and debt, suggestive of myopic behaviour. However, the work cited above is not based on panel data, and as noted may suffer from endogeneity.

### 6.4.2 New hypothesis on relative income effects and over-consumption

The BHPS is used here to explore whether relative income effects generate pressures on individuals and households that result in myopic behaviour in terms of excessive consumption revealed through low savings and debt problems.

**Hypothesis i:** Controlling for income, reference income will have a negative effect on savings, against a null of non-significance or positive effect.

**Hypothesis ii:** Controlling for income, reference income will have a positive effect on household debt, against a null of non-significance or negative effect.

### 6.4.3 Testing the new hypothesis on the BHPS

In each wave respondents are asked how much they save each month. Table 34 considers whether relative income has a negative impact on savings when controlling for own income, household income and lagged income. A Tobit model is used since many households have no monthly savings resulting in over 40,000 left censored observations.\(^{71}\) Only household heads are included here due to the problem of identifying individual and household debts. Additional controls of house value and housing costs are included since these may influence saving behaviour.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log own hh income (net)</td>
<td>105.01(3.069)**</td>
<td>104.82(3.079)**</td>
<td>104.91(4.826)**</td>
</tr>
<tr>
<td>Log own income (gross)</td>
<td>28.79(2.167)**</td>
<td>28.56(2.168)*</td>
<td>31.85(3.316)**</td>
</tr>
<tr>
<td>Mean hh income by region (net)</td>
<td>20.36(9.125*)</td>
<td>18.91(5.378)**</td>
<td></td>
</tr>
<tr>
<td>Mean hh income by region,</td>
<td>20.36(9.125*)</td>
<td>18.91(5.378)**</td>
<td></td>
</tr>
<tr>
<td>education and age (net)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean hh consumption by region</td>
<td></td>
<td></td>
<td>-5.56(14.387)</td>
</tr>
<tr>
<td>House value</td>
<td>2.282(0.2261)**</td>
<td>2.275(.2263)**</td>
<td>3.525(0.393)**</td>
</tr>
</tbody>
</table>

\(^{71}\) Under such circumstances OLS leads to bias estimates.
<table>
<thead>
<tr>
<th>Net housing costs</th>
<th>-0.965(0.0079)**</th>
<th>-0.965(0.0079)**</th>
<th>-0.074(0.0127)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>64,803</td>
<td>64,707</td>
<td>28,238</td>
</tr>
<tr>
<td>Number left censored</td>
<td>40,475</td>
<td>40,407</td>
<td>17,587</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-181381.98</td>
<td>-181164.32</td>
<td>-8014.806</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.0297</td>
<td>0.0297</td>
<td>0.274</td>
</tr>
</tbody>
</table>

Note: Waves 5-14 pooled, except consumption which is 7-14 due to data availability. Coefficient(se).

Savings appear to be enhanced by living in a region with higher average household income. When reference income is taken as the average of the age (four groups) and education (six groups) for each region this makes little difference to the findings. In both these cases the reference is average income rather than average consumption. Yet it is theorised that the consumption of others, especially conspicuous consumption, drives myopic consumption behaviour. However, when reference consumption is used (column three), although negatively related to savings, this is poorly defined and not significant.

Other determinants in the model support the reliability of the savings data. For example, savings are significantly increased by education, excellent health, being self-employed, and living in a metropolitan area. They are significantly reduced by being divorced, separated or cohabiting, being in fair or poor health, having children in the household of any age, being on long-term sick, retired, unemployed, on maternity leave, or being a family carer.

Waves 5, 10 and 15 contain detailed data on household assets and debts. However, looking at the extent of debt is problematic since debt is likely to be influenced by the availability of credit and investment opportunities. Declared household debt is positively correlated with household income for these waves (0.21), which suggests that debt may not reflect myopic behaviour. An alternative measure, which may come closer to reflecting myopic behaviour, is whether the existing level of debt creates problems for the household. From wave five respondents were asked, “To what extent is the repayment of such debts (hire purchase or loans) and the interest a financial burden on your household? Would you say it is a ...heavy burden, somewhat of a burden, not a problem.” Around 11% report a problem with debt repayments. It is therefore possible to consider whether relative income (controlling for own income) increases the likelihood of holding a debt burden that is a problem for the household.

Using a conditional logit fixed effects model (as outlined in Chapter Three) it is possible to consider those times when debt repayments create a heavy or somewhat of a burden on the household compared to those when it is not applicable or is not a burden. Table 35 shows these models, both for reference income as mean income for the region and mean income for the region, age and education group. This model only considers those who change their debt-burden status which dramatically reduces the sample size, consequently a random effects model is also
considered. Since the individual fixed effects are unlikely to be correlated with regional income, failure to control for fixed effects is unlikely to present too much of a problem for this coefficient.

### Table 35: FE Conditional logit and RE logit on 'debts being a problem'

<table>
<thead>
<tr>
<th></th>
<th>Regional mean as reference income</th>
<th>Regional mean by education group and age as reference income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(FE)</td>
<td>(RE)</td>
</tr>
<tr>
<td>Log own hh income (net)</td>
<td>.0310 (.0858)</td>
<td>-.0728 (.0704)</td>
</tr>
<tr>
<td>Log own income (gross)</td>
<td>-.0600 (.0435)</td>
<td>-.0902 (.0364)*</td>
</tr>
<tr>
<td>Regional mean hh income</td>
<td>-.9073 (.5333)</td>
<td>.8655 (.2842)**</td>
</tr>
<tr>
<td>Regional mean hh income for education &amp; age</td>
<td></td>
<td>-.1151 (.1929)</td>
</tr>
<tr>
<td>Lagged own hh income</td>
<td>.0219 (.6140)</td>
<td>-.0542 (.0517)</td>
</tr>
<tr>
<td>Standard controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>12,589</td>
<td>45,837</td>
</tr>
<tr>
<td>Individuals</td>
<td>2,016</td>
<td>9,167</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-4624.29</td>
<td>-12082.89</td>
</tr>
</tbody>
</table>

Note: Waves 6-14. Coefficient(se). Heads of households only.

Although the results are not conclusive, regional income does not seem to be increasing the risk of having problems with debts. Reference income is significantly negative within the random effects models, suggesting that, if anything, living in a higher income area reduces the problems of having difficulties with debt. This remains so when the reference income used is average regional income by education and age group (3rd and 4th columns).

Again the validity of the debt problem variable is supported by reasonableness of other determinants. In the fixed effects model, the likelihood of having a problem with debt is significantly increased for the self-employed, those aged 20-25 compared with being under 20, having teenagers in the house and having problems walking.

This analysis of the BHPS does not therefore find evidence to suggest that relative income or consumption effects result in over-consumption. However, this may be because the use of regional average income is too remote to the individual to impact strongly on behaviour.

### 6.5 Policy implications

#### 6.5.1 Policy implication of adaptation to income

There is some evidence that past income has a negative impact on SWB when controlling for current income, although lagged income coefficients are not always well defined within the
BHPS data. As yet there is no clear picture of the time scale, nor the processes involved behind this relationship. Increases in income are beneficial in terms of SWB, as are increasing trajectories, however, this type of analysis does not offer any evidence that individuals are not aware of this fact. Indeed, individuals are known to show a preference for such trajectories (e.g. Loewenstein and Sicherman, 1991).

If income taxation is proposed based on individual failure to anticipate adaptation to income it is necessary to know that adaptation to income will definitely occur and that it is not already factored into decisions (e.g. through the use of savings to maintain more constant consumption).

The BHPS analysis suggests that there is something more than adaptation to new levels of income driving the finding of a negative lagged income coefficient. If individuals become accustomed to a certain level of income we would expect past income to also have a negative impact upon income satisfaction and financial coping, however, this is not found to be the case. Past income effects may be picking up expectations about the future, but this does not fully explain the impact of past income. Another possible interpretation is that higher incomes are crowding out other beneficial goods.

The lack of robustness of the current income coefficient, and negative impact of past incomes suggests that any taxation on increases in income, particularly if they preserve a positive trajectory, are likely to result in a minimal loss in SWB in the long run. However, using the data to estimate the extent to which higher previous income reduces life satisfaction at current income levels is problematic due to the lack of robustness of the lagged income coefficients.

More research is required exploring the benefits of short and medium run problems arising from changes in income, both increases and decreases, and the potential changes in attitudes which follow increases in income in terms of increased attachment to material possessions.

### 6.5.2 Policy implications of relative income

The BHPS data suggest that relative income effects are robust to the inclusion of housing costs, and to lagged income. Evidence of negative externalities from others’ income implies that this is an essential input into deriving any appropriate taxation. Whilst this may be desirable, the findings from the literature and the BHPS suggest a lack of robustness of the relative income effect.

The collective action problem generated by the presence of relative income effects may already have been corrected for with existing labour income taxation. The finding that relative income effects are present does not therefore imply the necessity for incentives to reduce hours worked.
However, the existence of relative income effects should feed into any discussion on appropriateness of labour taxation policy, safety policy and maximum hours legislation.

In the BHPS relative income effects are confined to those who move regions. This may arise because attention in not drawn to the consumption behaviour of those in an area in which patterns of consumption of others change only gradually. The detrimental impact of increased relative income for movers suggests migrants into high-income areas, either within countries or internationally, may be particularly vulnerable to negative relative income effects.

Leibenstein (1962) noted that, “common emotions and considerations as envy, compassion, and relative status and so on do determine in part the utility that we derive from the outcome of economic events” (Leibenstein, 1962: 301). Our compassion leads us to benefit from others wealth, our envy and desire for relative status leads us to suffer. However, others’ income, particularly geographically based others, impacts in more ways than compassion, envy and status. Our neighbours may be ‘positives’ if they bring neighbourhood level status, reduced anti-social behaviour, or improved access to services. Since interdependencies can operate at many different levels and in different directions, and existing empirical work does not present a consistent verdict on the magnitude of the effect of reference income nor the appropriate reference group, it is hard to see reference income effects being sufficiently well defined to currently feed directly into taxation policy.

The impact of income inequality, controlling for own income, is heterogeneous. However, given that SWB only measures prudential value, and equality of resources is a moral issue as well as one of efficiency, this finding may have minimal relevance for policy. Furthermore, since there are diminishing returns to income, for any given level of regional income a more equal distribution of income enhances aggregate well-being.

**6.5.3 Policy implications from relative income effects and over-consumption**

Where consumption has status benefits, one individual’s consumption will impose a negative externality upon others. Taxation on consumption may be efficient if it is able to internalise the cost of these status losses to others. However, Veblen (1899) style ‘conspicuous consumption’ may be inelastic, indeed a high price of goods may attract more potential status from their consumption (Leibenstein, 1950). Taxation on specific goods in this case may not help establish the welfare maximising consumption level, but it does offer a potential source of revenue with few distortions (which can be directed towards goods whose consumption is not a zero sum game). However, if the negative externality of consumption is imposed upon those of below
average (or peer group) income and encourages excessive consumption and myopic expenditure behaviour, then consumption taxation which does not alter behaviour may aggravate the negative impact of others consumption. The assumption that taxation will reduce consumption implies a degree of rationality in the consumption decision that may not be present.

The BHPS provides no evidence that reference income at the regional level generates a reduction in savings, nor higher risk of experiencing problems with debt. Indeed the opposite may be the case, controlling for own income, low-income areas appear to experience less saving and greater difficulties arising from household debt.

Of course a reduction in consumption may have other benefits, and, importantly, reduce other negative externalities such as environmental damage. However, any policy recommendations based on the notion that relative income effects result in over-consumption require more evidence.

6.6 Conclusion

The BHPS data suggests that income has a small beneficial impact on SWB. Part of the beneficial impact of income is temporary, and is likely to have negative consequences for future SWB, particularly if future incomes are lower. Part arises because of changes in relative income position hence on aggregate across the reference group represents a zero sum.

A greater understanding of relative income effects is still needed, particularly understanding who comprises an individual’s reference group. Diener and Fujita (1997) noted that there was a lack of clarity in how income comparisons work, or exactly who individuals compare themselves to. Despite a growing amount of empirical work over the last decade, this is still the case.

Although concerns remain over the exact coefficients on reference income or past incomes, that they impact upon SWB in some way is reasonably well supported and replicated within the BHPS. In a dynamic framework where the individual lives within a context of others, income is clearly a poor generator of SWB, which further weakens any claim on income as a useful measure of well-being. Clear policy implications arise from the inadequacy of income to represent individual well-being. At a macro level, this casts doubt on the rationale for prioritizing economic growth. And at a micro level it implies the discontinuation of using WTP as a measure of value. If revealed WTP does not represent a change in well-being, this undermines its use as a building block of economic theory, and as a valuation for non-market goods.
A greater understanding of the complex relationships between current, past and others income and SWB is needed. More research is needed specifically to understand when income matters most. For example, income may be particularly influential to SWB in situations of poor health or difficult social circumstances or when future security is threatened.\textsuperscript{72}

Greater understanding of the situations in which income levels interact with individual capabilities may help deepen the understanding of the income-SWB relationship. For example, at what level of income do individuals suffer from low levels of control? Or more specifically, what combinations of income, debt, access to credit and expenditure demands result in low levels of control?

This thesis has given accumulating evidence that income is a very poor measure of well-being. The main failings of income rest on the fact that it is not a sufficient statistic for well-being. The impact of income upon well-being will depend, amongst other things, upon previous incomes over the recent or distant past and upon critical others’ incomes.

\textsuperscript{72} Older people have been shown to have fairly high levels of financial satisfaction (Easterlin, 2006), however, future security is also considered by older people to be an important attribute of well-being (Grewal \textit{et al.}, 2006).
APPENDIX 6.A

Externalities from adaptation to income

Table 36 shows an example of alternative working hours under considerable adaptation to income. In the first period the individual has an incentive to work additional hours if she does not anticipate adaptation or focuses only on the costs and benefits in the current time period. She may choose to work 45 hours giving a net gain of 2 units, as oppose to 1 unit at the original 40 hours. However, if the utility from additional income is only temporary, with time, much of this additional utility will fade away, yet working hours will still be longer. This is reflected in a decline in utility units to 9 in period two, and 8.5 in period three. From period three she would have been better off had she remained at 40 hours, however, taking a reduction in hours from her new position of 45 hours, and consequently a reduction in income from her current perspective requires taking a drop in utility (for simplicity this is the same magnitude as the initial rise, although a more realistic example would account for both diminishing returns to income and the asymmetry between gains and losses in income). Faced with a choice at period four of remaining at 45 hours or dropping down to 40 hours, she is better off in the short-run remaining at 45 hours. Moving only appears to the best option if she takes account of the future in which she re-adapts to lower income. In a world of unanticipated adaptation or myopic decisions work effort may extend beyond long-term interests.

Table 36: Hypothetical example showing incentives to change work hours

<table>
<thead>
<tr>
<th></th>
<th>45 hours</th>
<th>40 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disutility of work</td>
<td>Utility from income</td>
</tr>
<tr>
<td>Period 1</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Period 2</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Period 3</td>
<td>8</td>
<td>8.5</td>
</tr>
<tr>
<td>Period 4</td>
<td>8</td>
<td>8.5</td>
</tr>
<tr>
<td>Period 5</td>
<td>8</td>
<td>8.5</td>
</tr>
<tr>
<td>Period 6</td>
<td>8</td>
<td>8.5</td>
</tr>
</tbody>
</table>

In this example taxation on additional income such that there was no initial incentive to work 45 hours, or maximum hours legislation, would result in the optimal outcome.

Externalities from relative income

---

73 The benefits and costs arising from any individual’s work commitments may also be felt by others via changes in productivity, taxation, time spent with the family etc.
Using the example from Table 37, and leaving aside the issue of adaptation, the choice facing individuals A and B is between 40 hours work and 45 hours work. Disutility of work is 7 for 40 hours and 8 for 45 hours work, whereas because of relative income effects, benefit depends upon the other individual’s choice of work hours. Were individual A to choose to work only 40 hours, working 45 hours for individual B brings benefit of 10, giving a net gain of 2, however, if individual A were to also choose 45 hours the net benefit to both is zero.

**Table 37: Prisoner’s dilemma of work hours under the negative externality of relative income**

<table>
<thead>
<tr>
<th></th>
<th>Person B works 45hrs</th>
<th>Person B works 40 hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person A works 45hrs</td>
<td>Both [disutility = 8, utility = 8, net gain = 0]</td>
<td>Person A [disutility = 7, utility = 6, net gain = -1] Person B [disutility = 8, utility = 10, net gain = 2]</td>
</tr>
<tr>
<td>Person A works 40hrs</td>
<td>Person A [disutility = 7, utility = 6, net gain of -1] Person B [disutility = 8, utility = 10, net gain of 2]</td>
<td>Both [disutility = 7, utility = 8, net gain = 1]</td>
</tr>
</tbody>
</table>

With co-ordination both individuals are better off working 40 hours, however, without co-ordination the potential benefit from a higher relative income position leads to them both working 45 hours. In this case a tax to reduce the utility from the 45-hours option, or maximum hours legislation, would ensure a cooperative outcome.
Chapter 7: Does taking to our neighbours enhance our satisfaction with life?

7.1 Introduction

Chapter Three showed that frequent contact with friends, family and neighbours increases well-being as measured by the SWB measures, yet has a negative relationship with well-being when measured by income and consumption. This is therefore an area where SWB measures can offer new understanding. This chapter will explore whether this new understanding leads to any clear policy recommendations.

A substantial body of researchers have drawn links between social capital and both collective and individual well-being. For example, social capital, perceived as a form of social trust and networks, has been linked to “lower crime rates, improved child welfare, better public health, more effective government administration, reduced political corruption and tax evasion, and improved market performance, education performance etc” (Helliwell and Putnam, 2004: 1437). Whilst the link from social capital to happiness may be mediated via such factors as health, employment, and reduced crime, there may also be a direct route between social capital and happiness, for example, Bjørnskov claims that “social capital is a powerful factor when explaining why some nations are happier than others” (Bjørnskov, 2003: 3). Helliwell and Putnam claim that, “social capital is strongly linked to subjective well-being through many independent channels and in several different forms.” (Helliwell and Putnam, 2005: 455).

Having empirical evidence of the link between social capital and SWB is important from a policy perspective since it may indicate the importance of sources of well-being which have been overlooked in existing policy. However, before this relationship can lead to any policy recommendations to enhance social capital, it needs overcome various methodological hurdles.

Firstly, clarity is needed on which specific aspects of social capital are related to SWB. ‘Social capital’ is a somewhat nebulous concept, defined in slightly different ways by different researchers, and measured in different ways (for example, using generalised trust, volunteering, membership of associations etc.). It is not clear if these are all proxies for the same underlying concept or slightly different concepts. Woolcock (1998) argues that vagueness has plagued social capital scholarship and as a single term it is “inadequate to explain the range of empirical situations demanded of it, that it confuses sources with consequences, justifies contradictory social policies, and understates corresponding negative aspects” (Woolcock, 1998: 159).

Different aspects of social capital may have different relationships with SWB and, as noted by
Woolcock, some aspects in certain situations may even be detrimental to individual well-being. For example, close connections within a group of people (integration or bonding capital), may weaken connections across different groups of people (linking or bridging capital) or infringe upon individual autonomy and personal freedoms. Policy implications cannot be drawn without a good understanding of the different aspects of social capital and how they individually and collectively impact upon well-being. In order to generate empirical evidence to support policy initiatives it is necessary to focus on measurable aspects and explore their impact.

The social capital literature has already moved towards presenting social capital as a combination of separate dimensions. For example, Bjørnskov (2006) uses principal components analysis on three waves of the WVS to show that social capital is comprised of three distinct features; generalised trust, average norms and associational activity. This chapter considers just one specific aspect of social capital - the frequency of communication with neighbours, and its relationship to SWB. This is both non-ambiguous and easily measurable.

Secondly, empirical work must show that any relationship between social capital and SWB is not spurious. For example, SWB is known to fluctuate with age, hence if age is not controlled for any association between social capital and SWB may be misleading. Similarly, there may be a link between personality and both SWB and social capital. Those with a happy personality may both have more friends and higher SWB. Studies require either that personality is somehow controlled for, or they require access to panel data which enables controls for unobserved personality differences - either looking at year on year changes (first differences) or using within person deviation from the individual mean (fixed effects).

Thirdly, empirical work must show that the direction of causality runs from social capital to SWB. Policy recommendations drawn from studies showing strong association between social capital and SWB may mislead if that association arises because of the impact which SWB has upon an individual’s social capital. Without evidence of a causal relationship recommendations for the extension of government organised volunteering and community engagement based on the premise that this will enhance well-being (e.g. NEF, 2004) may be premature.

Since it is reasonable to assume that when people feel happy they are more sociable, reverse causality presents a particularly thorny problem. Whilst the link between social capital and SWB is fairly well established in the literature, there is little evidence that withstands the criticism that it may be driven by personality differences or reversed causality. To address this, additional analysis is conducted on the BHPS data which attempts to address issues of potential endogeneity in the social capital variable.
Chapter Three showed that having frequent contact with neighbours was significantly positively associated to all SWB measures within the BHPS. The coefficients reduce slightly in fixed effects, but contact with neighbours is one of the few variables that is sufficiently robust to show positive effect both within and between people, and across the range SWB measures.

This chapter seeks to explore this relationship in more detail. Fully understanding this relationship is important for two main reasons. Firstly, because this is an area where recommendations would conflict depending on whether subjective or income/consumption based well-being measures were used (high contact with neighbours being negatively associated with income). Secondly, because this is an attractive area for policy recommendations in which there may be potentially cost effective policy interventions that could enhance individual well-being.

7.2 Background to social capital and subjective well-being

Existing literature has explored the impact of social capital measured in various ways, using both cross-country and within-country analysis. For example, Bjørnskov (2003) finds an index of social capital (comprised from responses to questions about trust, a measure of civic participation, and perceptions of corruption) to be positively associated with life satisfaction using international data. He shows that a change in social capital score of 10% of the distance between the highest and lowest score results in an increase in life satisfaction of 4.5%.

However, presenting an index of social capital clouds the picture of exactly which attributes are enhancing well-being, and a more constructive approach is to consider the separate contribution of different aspects of social capital. The following sections summarise current evidence on the association between SWB and membership of organisations, volunteering, generalised trust, contact with friends and family, and contact with neighbours.

7.2.1 Membership of organisations

A positive relationship is found in some studies between SWB and membership in organisations (Gundelach and Kreiner, 2004; Helliwell, 2003; Helliwell and Putnam, 2004; Pichler, 2006). However, contrasting results are found by Li et al. (2005) who find that civic participation has a negative but non-significant relationship on life satisfaction in the BHPS and Bjørnskov (2006) who finds that when trust and social norms are controlled for associational activity in the WVS has a significantly negative effect on life satisfaction. However, this effect disappears when regional dummies are left out of the equation.
Helliwell and Putnam note that some of the effects of association membership may lead to higher levels of trust, hence a stronger impact of membership on life satisfaction would be anticipated in models which do not control for trust (Helliwell and Putnam, 2004: 1443). This may explain the findings by Bjørnskov (2006), however, there is no evidence to suggest that the impact of trust on SWB is mediated via organisational membership.

7.2.2 Volunteering

The evidence on the link between volunteering and SWB is also mixed. Putnam (2000) finds a positive relationship between happiness and volunteering in the US DDB Needham Life Style Survey. However, using cross-country data (WVS, 1995-97) Haller and Hadler (2006) find no such relationship. Differences may arise if volunteering impacts upon some aspects of SWB and not others and SWB is measured differently in different studies. For example, Greenfield and Marks (2004) find that volunteering in older people is associated with more positive affect, more meaning in life, but not less negative affect.

This difference may also arise if volunteering has a different impact in the US than elsewhere. However, a positive impact of volunteering has also been found in Germany (Winkelmann, 2006). Meier and Stutzer (forthcoming) take advantage of data from Eastern Germany where institutional breakdown following the collapse of the German Democratic Republic resulted in a dramatic reduction in volunteering. They find that more regular volunteering increases life satisfaction. The impact of volunteering reduces considerably when individual fixed effects are controlled for, suggesting that part of the higher well-being levels may arise from individual personality differences (Meier and Stutzer, forthcoming).

Analysis of data from ACL survey, finds that although measures of personal well-being have a positive relationship with volunteering, it is also the case that happier people tend to do more voluntary work (Thoits and Hewitt, 2001). This draws into question the direction of causality in this relationship.

7.2.3 Trust

Trust can be conceptualised at four levels, generalised or social trust, which relates to trust in strangers; trust in people one knows; trust in institutions and how trustworthy one perceives oneself to be. At all four levels, evidence suggests trust is beneficial for well-being.

Using the WVS and the ESS data, Helliwell (2003; 2006; Helliwell and Putnam, 2004) finds that social trust is associated with higher life satisfaction and happiness, and a lower probability of suicide. For the UK, Li et al. (2005) use cross-section analysis on the BHPS and find
neighbourhood social trust increases life satisfaction. Trust in key public institutions such as the police, the legal system and government is also associated with higher life satisfaction (Helliwell and Putnam, 2004; Hudson, 2006), as are beliefs about the wrongness to cheat on one's taxes (Helliwell, 2003).

Yip et al. (2007) find a positive impact of trust on life satisfaction and mental health in a cross-section survey in rural China, with high levels of trust increasing the probability of collective action occurring and more emotional support being available.

In analysis of the WVS, Bjørnskov (2006) finds a strong positive effect on SWB of both a generalised trust component derived from principal components analysis, and a social trust measure as a simple percentage of the population with positive response to the question “In general, do you think that most people can be trusted, or can’t you be too careful”. As noted above, this study also finds a negative impact of associational activity, suggesting the effects of social capital are ‘driven entirely by social trust’ (Bjørnskov, 2006: 33).

### 7.2.4 Contact with friends and family

Socialising with family and friends tends to be fairly unambiguously positively associated with SWB (Diener and Seligman, 2004; Dockery, 2003; Helliwell and Putnam, 2004; Lelkes, 2006; Li et al., 2005; Magdol, 2002; Martin and Westerhof, 2003; Powdthavee, forthcoming; Taylor et al., 2001; Winkelmann, 2006). The quality of social networks is also found to be related to well-being in older people (Baker et al., 2005; Ritchey et al., 2001; Wiggins et al., 2004) as is the number of people who can provide social support (Baker et al., 2005). Despite strong association based evidence, there is little evidence that close contact with friends and family is the cause of higher SWB. In the only study to control for individual personality, and address the issue of reverse causality, Powdthavee (2005) found evidence from the BHPS of a causal relationship from seeing friends and family to enhanced life satisfaction. This study instruments for contact with friends and family using the price of fuel, on the basis that contact will be reduced when traveling costs are higher.

### 7.2.5 Contact with neighbours

In Bowling Alone Putnam argues that for the US socialising with neighbours is linked to general levels of happiness (Putnam, 2000). Later cross-section analysis conducted with Helliwell supports this finding using the US Benchmark 2000 data and the Canadian ESC 2000-2003 data, although this reaches significance only in the Canadian study (Helliwell and Putnam, 2004). Analysis of the BHPS by Li et al. (2005) shows that neighbourhood attachment (a latent
variable based feelings and behaviour related to the neighbourhood, including talking with neighbours) increases life satisfaction, improves health and reduces the risk of depression, even when controlling for trust, social networks and civic participation. Powdthavee finds a strong relationship between talking to neighbours and life satisfaction also using the BHPS (Powdthavee, *forthcoming*). This effect reduces in magnitude when controlling for individual fixed effects but remains significant.

Of all the social capital variables, contact with neighbours is especially vulnerable to endogeneity problems. If individuals are currently unhappy they may be unlikely to initiate contact with those around them. Low current mood is likely to be less of a deterrent to frequent contact with family and friends than with neighbours. Family and friends are more likely to maintain contact during periods when the individual is feeling less sociable, partly because of the caring role friends and family may adopt and partly because social ties and bonds are likely to be stronger than those with neighbours.

### 7.3 Data and methodology

Since wave 7 of the BHPS respondents have been asked how often they talk to their neighbours; “*How often do you talk to any of your neighbours? Is it ... most days / once or twice a week / once or twice a month / less than once a month / never*”. The data\(^{74}\) show a clear bivariate relationship between life satisfaction and the frequency of talking to neighbours. As a simplification the frequency of talking to neighbours is coded onto a 1-4 ‘talking to neighbours scale’. Since never talking to neighbours and talking less than monthly have a similar impact on life satisfaction and there are few observations in the last category these are combined. Never talking to neighbours or talking less than monthly is given a score of 1, talking to neighbours monthly is given a score of 2, weekly 3 and most days 4. Figure 14 shows that the average of this scale is associated with higher life satisfaction responses. Similarly, Figure 15 shows the distribution of talking to neighbours, and that mean life satisfaction increases with the frequency of talking to neighbours.

\(^{74}\) This sample excludes students and those from Northern Ireland, both groups were considered to have different relationships towards their neighbours.
Does talking to neighbours enhance satisfaction with life?

Figure 14: Frequency of satisfaction with life, BHPS wave 14, with mean (sd) of talking to neighbours scale (1-4)

![Bar chart showing frequency of satisfaction with life, BHPS wave 14, with mean (sd) of talking to neighbours scale (1-4).]

Figure 15: Frequency of talking to neighbours, BHPS wave 14, with mean (sd) life satisfaction (1-7 scale)

![Bar chart showing frequency of talking to neighbours, BHPS wave 14, with mean (sd) life satisfaction (1-7 scale).]

However, as discussed above, this relationship could be spurious. Possible variables that could be related to both neighbourhood contact and SWB are employment status, hours worked, an
individual’s health, neighbourhood harmony, family structure and personality. The following considers each of these in terms of their possible connection to both variables and how they are controlled for within the subsequent analysis.\footnote{See appendix 3.D for a full description of the variables used.}

The frequency of talking to neighbours may be related to the number of hours worked. Working more hours may mean less time available to chat to neighbours. Although being unemployed is clearly detrimental to SWB (see Dolan \textit{et al.}, 2006a for a review), working for a high number of hours may also be detrimental (Luttmer, 2005; Meier and Stutzer, 2006). Work commitments are controlled for here by employment status (being employed compared with being self-employed, retired, unemployed, on long-term sick leave, a home-maker, or other job status) and the total hours worked (including basic and overtime hours, commuting time, housework, and caring work). A squared term is also included to allow for a non-linear relationship between hours worked and life satisfaction.

Health status may be related to the frequency of talking to neighbours, either negatively if those with restricted mobility or anxiety/depression have less desire or opportunity for involvement; or positively if those with health problems spend more time at home and in the neighbourhood. Studies find physical and psychological health states to be strongly related to SWB (e.g. Clark and Oswald, 2002; Easterlin, 2003; Diener and Seligman, 2004; Shields and Wheatley Price, 2005). Fortunately, the BHPS contains a wealth of health information, enabling careful control for health status. The controls used here are a subjective health scale, whether the respondent has problems walking, whether they are registered disabled, and the number of inpatient hospital days (excluding births). The assumption is made that the health variables can be treated as exogenous.

The degree of peace and security in the neighbourhood could affect frequency of contact with neighbours (either positively because of shared desires to address common problems, or negatively due to fear and lack of trust generated by disruptive behaviour) and SWB directly (for example, through lack of sleep). There has been little empirical work estimating the loss in SWB from social disorder and disruption. Social harmony is controlled for by including responses to questions on crime and vandalism, noise on the street, and noise from neighbours.\footnote{From wave 6 onwards respondents were asked \textit{“Does your accommodation have any of the following problems…noise from neighbours; other street noise; vandalism or crime in the area”}.}

Family structure, marriage status, and the presence of children may increase or decrease the likelihood of involvement with neighbours. Households composed of large extended families may be more self-contained, therefore have less need to seek involvement with neighbours,
alternatively, the presence of children may lead to interaction with other households with children living nearby. Since marriage status and family are important to SWB (recent studies showing this include Helliwell, 2003; Lucas, 2005; Shields and Wheatley Price, 2005) controlling for them is essential. Family relationships are controlled for here through marriage status (being married compared to being divorced or separated, cohabiting, never married, and widowed), the frequency of seeing friends and family (most days compared to weekly, monthly, less than monthly, and never) and the number of children and adults living in the household.

Age or cohort effects may also be related to talking to neighbours and show a strong relationship with SWB. The relationship between age and SWB has generally been found to be ‘U’ shaped (e.g. Blanchflower and Oswald, 2004). Mean life satisfaction by age in the BHPS suggests an ‘S’ shape. Age is included here in 5 yearly categories the reference category being 20 or below.

The length of time lived in the neighbourhood may also be related to both talking to neighbours and SWB. For example, those who have recently moved may experience stresses related to moving home and job, and have little knowledge about accessing services in their new area. This is controlled for by the number of years an individual has lived in their current house. This variable is highly correlated with age (0.57).

Income is controlled for via household income (log of real, current, monthly, net income) and the individual’s perceptions of whether they are coping financially (‘living comfortably’ compared with ‘doing alright’, ‘just about getting by’, ‘finding it quite difficult’ and ‘finding it very difficult’. This subjective assessment of income should help pick up both the impact of past income and reference income upon life satisfaction, as discussed in Chapter Six. Gender, education level (no qualifications compared with O-level or similar, A level or similar, and higher education), and whether the individual is a high-hours carer are also included as controls as are time and regional dummies along with a dummy variable for whether the area is metropolitan (the variable ‘urban’).

An individual’s personality, particularly qualities such as extraversion, is likely to be positively related to talking to neighbours. Extraversion is also known to be related to SWB (DeNeve and

77 The inclusion of both age and wave dummies is problematic in fixed effects models using balanced panels since in any wave each individual will be an identical amount above their individual mean age. Although in an unbalanced panel, as is used here, this presents less of a problem the use of age groups is preferred for clarity.

78 This variable contains many missing values for Scotland, Northern Ireland and Wales, giving non-representative average values for these areas.
Chapter 7  Does talking to neighbours enhance satisfaction with life?

Cooper, 1998). Since there are no personality variables available in the dataset,\textsuperscript{79} this must be controlled for in other ways. With access to panel data, individual personality can be controlled for by comparing the relationship between an individual’s circumstances and their SWB at different periods of time - the within-person comparison. The individual effect therefore shifts the level of SWB; a happy disposition for example, or tendency to reply to survey questions in a positive manner, would shift the level of SWB upwards in each time period.

Within-person analysis is more easily conducted with continuous variables using ordinary least squared (OLS). As discussed in Chapter Three treating the life satisfaction variable as continuous has been shown to make little difference to the results compared with models that treat the life satisfaction variable as ordinal. Given the advantage of the OLS methods in terms of interpreting the coefficients, and easily controlling for the individual heterogeneity, fixed effects OLS will be used here.

The regressions here adopt the following form where life satisfaction is a function of speaking to neighbours (\(N_{it}\)), a vector of other variables (\(Z_{it}\)), an individual effect (\(\alpha_i\)) and an error term (\(\varepsilon_{it}\))\textsuperscript{80}

\[
LS_{it} = \beta_1 N_{it} + \beta_2 Z_{it} + \alpha_i + \varepsilon_{it} \tag{1}
\]

To get estimates of the coefficients it is necessary either to estimate the unobserved, individual effect\textsuperscript{81} (\(\alpha_i\)) or remove it from the model. A fixed effects approach removes the individual effect from the model by using the deviation from the individual mean for each variable (as discussed in Chapter Three). However, fixed effects are not able to control for time variant unobserved factors, such as time variant reporting and the impact of current mood. For example,

\textsuperscript{79} Personality variables are included in wave 15.
\textsuperscript{80} where \(\varepsilon_{it}\) is assumed to be independent and identically distributed over individuals and time with mean zero and variance \(\sigma^2\).
\textsuperscript{81} This can be done using a ‘Mundlak’ approach (Mundlak, 1978; Hsaio, 2003). The assumption \(E(\alpha, X_i) = 0\) is relaxed by assuming that the individual effects (\(\alpha_i\)’s) are linearly related to the independent variables (\(X_i\)’s): \(\alpha_i = a X_i + v_i, v_i \rightarrow N(0, \sigma^2)\). Hence the aXi picks up the correlation between the fixed, unobservables and the observable independent variables, which gives:

\[
LS_{it} = \beta X_{it} + a X_i + v_i + u_{it} \rightarrow E(\nu_i) = E(u) = 0
\]

GLS applied to this produces unbiased and efficient estimators of \(\beta\) and a. The coefficient on \(a\) can be interpreted as a correction factor, as done by Ferrer-i-Carbonell (2005). However, Mundlak claims that the \(a\)’s have no meaningful interpretation. Whether this approach adequately controls for individual heterogeneity depends on the appropriateness of modelling the individual effects as a function of the explanatory variables. Although experimenting with these models within an ordinal framework, the uncertainty surrounding their interpretation and the limited sacrifice of interpreting the life satisfaction variable as cardinal, led to a focus on fixed effects as a means of addressing unobserved heterogeneity.
a good mood may lead an individual to overstate both their life satisfaction and the frequency of talking to their neighbours. If that is the case then measurement error on the two variables will be correlated.

People with bipolar mood disorder tend to avoid social contact or new experiences when depressed which would suggest a direction of causality, at least for some individuals, running from mood and happiness to sociable behaviour (Lykken and Tellegen, 1996; Lyubomirsky et al., 2005; Diener and Seligman, 2004). Consequently, an endogeneity problem exists, and the variable talking to neighbours may be correlated with the error term resulting in the coefficient ($\beta_1$) being bias even in very large samples. To gain a consistent estimate of $\beta_1$, it is necessary to strip away correlation with the error term, but leave that part of the ‘talking to neighbours’ variable that is correlated with SWB.

In theory, this can be done using instrumental variables (IV). Rather than including the frequency of talking to neighbours instruments are found which correlate with talking to neighbours, but not the error term. This involves predicting the frequency of talking to ones neighbours through variables that are not related to SWB. A positive relationship on the instrumented variable would be indicative of a causal relationship running from talking to neighbours to life satisfaction.

### 7.4 Results

Talking to neighbours is found to have a positive and significant relationship with life satisfaction, and this relationship is monotonic (Model 1, Table 38 in Appendix 7.A). Compared to talking to neighbours monthly, talking most days raises life satisfaction by 0.051 points (on a 1 to 7 scale), or about one third of a standard deviation. As a further simplification (Model 2) the ‘talking to the neighbours scale’, is treated as a continuous variable.

These results are generally consistent with the understanding of the determinants of SWB from the literature (see Dolan et al., 2006a). Controlling for the perception of financial situation leads to actual household income being insignificant. The relative importance of perceptions of financial situation may arise either if perceptions are a more accurate measure of real income (accounting for local prices) or if perceptions and attitudes towards circumstances are more important than the reality of those circumstances. Job status impacts in the usual manner. A useful distinction is made here between long-term sick leave, and not working, the former having a greater reduction in life satisfaction than being unemployed. Time spent working has a significant effect and follows a U shape, maximising life satisfaction at around 35 hours.
Chapter 7

Does talking to neighbours enhance satisfaction with life?

Less than 7% of the variance in life satisfaction is explained by within-person changes in these variables. Whilst this is small, it compares favourably with other studies using fixed effects models on SWB.

To test whether ‘talking to neighbours’ can appropriately be treated as exogeneous requires finding instruments which are (correctly) not in the original model and are highly correlated with ‘talking to neighbours’. A good starting place in the search for instruments is considering those variables that are associated with the frequency of talking to neighbours. For example, a study in Ireland found that the extent of knowing neighbours was related to the ‘walkability’ of the neighbourhood, age, presence of children at home, attending religious services and the number of years lived in the neighbourhood (Leyden, 2003). A regression is run on the ‘talking to neighbours scale’, which includes variables from the regression on life satisfaction and others thought to potentially influence the extent of contact with neighbours. In addition to helping to identify possible instruments, this exploration will also shed light on whether the frequency of talking to neighbours is likely to be something that can be influenced by government intervention.

Table 39 (Appendix 7.A) presents both fixed effects and random effects regressions on the frequency of ‘talking to neighbours’. Importantly, life satisfaction in the previous period is positively related to talking to neighbours, implying a direction of causality running from SWB to being sociable with neighbours.

Household income is negatively related to ‘talking to neighbours’, significantly so in the random effects model, suggesting those with higher income have less contact with neighbours. Married people have more contact with neighbours than divorced or separated, or cohabiting individuals. Respondents who note a problem with noisy neighbours or noise on the street have less contact with neighbours, suggesting that the absence of civil harmony reduces social connectedness. The health variables show a negative relationship, the disabled and those reporting poor or very poor health status having significantly lower contact in both the fixed and random effects models.

All age groups report greater contact with neighbours than the 20s and under group, this relationship increases with age until the 66-70 age group, after which it decreases. This suggests there may be some cohort effects operating where older individuals have more contact with neighbours, but beyond a certain point age reduces contact.

Number of adults in the household follows an inverse U relationship, initially increasing then decreasing the frequency of contact (being at its maximum at 3 adults). As expected, the number
of children, particularly young children increases contact with neighbours. Those who stay at home have more contact with neighbours, whether retired, a home carer or unemployed. Women have significantly more contact than men.

Time spent working is not significantly related to talking to neighbours, however, the number of years lived at the address is significantly negatively associated. This is surprising and warrants future investigation. Years spent at the address is related to other variables within the model including age, education status, number of children in the household, living in an urban area, marriage status, gender and housing tenure, hence it may be picking up other effects.

Type of housing is significantly related to contact with neighbours. For example, living in a terraced house increases the 1-4 scale by 0.0838 in fixed effects (0.14 of the within person standard deviation) or 0.1812 in random effects (20% of a standard deviation). This suggests that the physical layout of housing influences the frequency of contact. Sharing a garden is also positively related to the frequency of talking to neighbours. Compared to home ownership, those who rent from the council or a housing association also have more contact with neighbours. However, those who rent privately have less. The district mean for talking to neighbours is significantly positively related in both models. This suggests that living in an area with more neighbourly contact encourages others to have more neighbourly contact.

The model in Table 39 presents a fairly detailed understanding of who has frequent contact with their neighbours. That this variable is significantly related to physical differences in house layout and relates to most variables in an anticipated manner offers additional confidence in the validity of this measure. Furthermore, the significance of the district mean of ‘talking to ones neighbours’ suggests an element of externality to this variable.

Returning to the search for instruments, one or more variables are sought which are correlated with the endogenous variable ‘talking to neighbours scale’, and are orthogonal to the error process in the life satisfaction regression. To make a valid instrument the relationship between the instrument and life satisfaction should be mediated via talking to neighbours. Clearly many variables above are already found within the life satisfaction model. House type has tended not to be included as a predictor of SWB, and although non-significant if included, and not found to be related to the error process, it is hard to have any confidence that these variables could not relate to SWB except via frequency of contact with neighbours, as such will not make intuitively appealing instruments.

---

82 This is estimated as a mean of all individuals within the local authority district, excluding the current individual. The BHPS is not ideally suited to aggregation at a district level; however, this was chosen in preference to regional averages since the aim is to consider the impact of neighbourhood connections in an individual’s local area.
The district mean of the ‘talking to neighbours scale’ is a potential candidate. If scial capital is taken as a public good, then this variable also arguably impacts directly on SWB. As noted earlier, Helliwell and Putnam’s analysis of the WVS finds that greater social participation by others increases SWB even for those whose own participation is not increased. However, for this data, the district mean of ‘talking to neighbours scale’ is appropriately excluded from the main model, and not related to the error term. Another useful instrument is the lagged value of talking to neighbours.

Model 5 in Table 38 shows that instrumenting for talking to neighbours using the district mean of talking to neighbours and the lagged value, gives a considerably higher coefficient, however, this is not well-defined, and is no longer significant. Interestingly, the coefficients on seeing friends and family reduce in size and lose their significance, suggesting that some of this benefit is being picked up in the instrumented ‘talking to neighbours scale.’

The fit of the first stage regression and explanatory power of the excluded instruments, suggests them to be fairly weak. Although the instruments are individually and jointly significant, and pass the rule of thumb that the F-statistic should be above 10 (Staiger and Stock, 1997: 557), these results are marginal (displayed at the bottom of Table 38). The Hansen J statistic (Wooldridge, 2002: 123) does not reject the null that all instruments are uncorrelated with the error term, suggesting the instruments are valid. However, Bound et al. (1995) show that if instruments are weak, even a small correlation between the error and the excluded instruments can lead to bias, even in large samples. The Davidson-MacKinnon test of exogeneity does not reject the null that the OLS-FE provides consistent and efficient estimates, which would suggest that talking to neighbours is not endogenous, or at least that there is little gain in terms of reduced bias from using an IV approach.

The magnitude of the IV coefficient suggests that a move from talking to neighbours most days to talking to neighbours monthly has a greater impact on life satisfaction than a move from

\[ \beta^{IV} = \beta^{OLS} + \frac{\text{plim Cov}(Z, \delta)}{\text{plim Cov}(Z, X)} \]

hence if the numerator is not 0, the smaller the value for Cov(Z, X), or weaker the relationship between the instrument and the endogenous variable, the greater the inconsistency. A weak relationship between the instrument and the endogenous variable also results in higher variance in the IV estimator than found in the OLS model. In large samples, the variance of the IV estimator can be evaluated as

\[ \text{Var}(\hat{\beta}^{IV}) = \text{Var}(\hat{\beta}^{OLS}) + \frac{1}{(\text{Cov}(Z, X))^2} \]

hence the closer to 1 is the sample correlation between X and Z the closer the variance of the IV estimator will be to the OLS estimator.

---

83 A regression with residuals from the model predicting life satisfaction as the dependent variable does not find the lag of talking to neighbours scale or mean level of talking to neighbours for the district to be significant.

84 However, excluding contact with friends and family from this regression does not substantially alter the coefficient on ‘talking to neighbours scale’.

85 This can be seen in the case with one instrument (Z) and one endogenous variable (X) where
employment into unemployment, and is nearly as damaging as a move from being married to being divorced or separated. This suggests either the importance of neighbourhood contact on individual well-being has been considerably underestimated, or the presence of bias leading to an over-inflated coefficient. The initial concern with endogeneity of talking to neighbours arose because of a possible unobserved variable, current mood, being positively related to both reported life satisfaction and talking to neighbours, and a direction of causality running from higher life satisfaction to increased contact with neighbours. These results, however, suggest that the endogeneity within the talking to neighbours caused an underestimate of the coefficient, for which it is more difficult to find a clear theoretical explanation.

7.5 Discussion

Talking to neighbours is positively related to life satisfaction, even when controlling for a wide range of social, economic and personal variables and individual time-invariant effects. Using an instrumental variable (IV) approach suggests that when the talking to neighbours variable is stripped of its correlation with the error process, a positive relationship is still found, yet this is no longer significant. The significant effect of the lag of life satisfaction upon talking to neighbours, and the weakness of the instruments used within the IV regression, denies any clear conclusion that the direction of causality is from talking to neighbours to life satisfaction.

The weakness of the instruments used here leads to a poorly determined coefficient, and results that may be bias. If suitable, strong instruments could be found both these problems may be solved. The use of politically defined district boundaries is far from ideal, and using the average of speaking to neighbours in much smaller geographical areas may be more appropriate. There may also be some potential in exploring other geographical features of neighbourhoods such as the ‘walkability’ of neighbourhoods. Unfortunately, this is not possible with the BHPS.

Some of the benefit arising from close relationships with neighbours may arise indirectly via the impact those relationships have upon characteristics which have been controlled for, such as income, employment, health and neighbourhood harmony. In terms of income, national economic performance has been linked to levels of trust and civic norms (Knack and Keefer, 1997; Zak and Knack, 2001), and expanded transaction opportunities via reduced transaction costs (Helliwell, 2001). Close community relationships may also reduce unemployment if personal connections lead to expanded job opportunities (Granovetter, 1983; Winkelmann, 2006). Community embeddedness has also been linked to improved health and mortality outcomes (Putnam, 2000; Kawachi et al., 2004; Rose, 2000). A close community may have advantages in solving collective action problems (Ostrom and Gardner, 1993; Putnam, 2000), such as addressing street noise, and lowering crime. For example, Freudenburg (1986) analysed
four communities in Western Colorado one of which had undergone rapid growth and found that high density of acquaintanceship results in better control of deviance and better socialization of the young. These indirect links would suggest that the benefits in terms of life satisfaction from regularly talking to neighbours, explored by controlling for income, health and social harmony, may be underestimated.

However, when satisfaction with finances is controlled for, talking to neighbours may be overestimated if those who regularly talk to their neighbours suffer more from detrimental social comparisons. It has been shown that the negative consequences of living in a high-income neighbourhood in the US are more intense for those people who socialise locally (Luttmer, 2005). Consequently, for those who live in high-income areas, frequent contact with neighbours may result in greater dissatisfaction with any given level of household income. However, this was not found to be the case in Germany by Knies et al. (2007) and further analysis which includes an interaction term for regional average income and neighbour contact (whether one sees ones neighbours on most days), also does not support the finding that greater neighbourhood contact results in a larger negative impact of neighbourhood income.86

It may also be the case that close contact with neighbours may be detrimental in circumstances where there is conflict between households. For example, those who suffer from noisy neighbours may not benefit from contact with their neighbours. Both ‘talking to neighbours’ and suffering from noisy neighbours are strongly related to house type. In wave 14, 2% of those living in a detached property report noisy neighbours, 9% living in semi-detached, 12% in terraced and 23% of those living in flats. Additional analysis on only those who suffer from noisy neighbours finds that talking to one’s neighbours is still positive but no longer significant.87

As discussed in the introduction there are advantages in focusing on one specific aspect of social capital, such as talking to neighbours, which is clearly identifiable. Talking to neighbours can be more easily grasped than the broader concept of social capital, and as such lends itself more easily to measurement, analysis and intervention. However, concepts such as social norms, reciprocity and trust are by their nature difficult to measure, but that does not preclude importance to well-being. Although effort has been made here to control for observable factors that may be driving the association between talking to neighbours and life satisfaction, it is still possible that unobservable aspects of social capital, such as sense of belonging or feelings of

86 In fact, those who see their neighbours most days have a smaller negative coefficient on regional mean income, although in neither case is this coefficient well defined.
87 However, the sample size is substantially reduced at only 5,000 observations, half of which are only observed for one period.
trust are driving the results. Furthermore, there may be considerable variability in what is meant by talking to neighbours, from a simple ‘hello’ in the street to a conversation with a close friend who lives nearby. These judgements may themselves be related to an individual’s SWB. Frequency of talking does not distinguish the quality of contact and this may have important implications for the impact upon life satisfaction.

7.6 Conclusion

A clear causal link from talking to neighbours to SWB has yet to be proved. However, the evidence of association is sufficient to justify further work to explore this relationship. Experimental evidence is required to complement analysis from large secondary datasets. Ideally evidence would be based on randomised controlled trials, where an intervention to enhance talking to neighbours is implemented and monitored across randomly allocated individuals. However, practical considerations may mean that group randomisation (e.g. whole areas) or before - after trial designs are more appropriate. Experimental work has the potential to answer a range of essential questions.

Firstly, is there a causal relationship between the frequency of talking to neighbours and life satisfaction? Secondly, can policy intervention change the frequency with which people communicate with their neighbours? Ideally, studies would also show the most cost-effective means of enhancing communication with neighbours. A secondary question arises as to whether intervention to enhance neighbourhood contact is a cost effective means of enhancing well-being compared to other possible government interventions. One of the biggest hurdles to this kind of research is likely to be requirements of sample size and subsequent cost implications. The impact of talking to neighbours is sufficiently strong to be identified with 10,000 cases, but may result in inconclusive results on scale which is practical for this kind of intervention study. One possibility may be to identify the attribute that mediates the relationship between talking to neighbours and life satisfaction, which may be satisfaction with a life domain such as satisfaction with one’s community. If it is possible to show a significant (causal) relationship between talking to neighbours and satisfaction with one’s community, and similarly, a significant (causal) relationship between satisfaction with one’s community and satisfaction with life then, this opens the potential for the outcome measure of experimental studies to be satisfaction with one’s community, which would be expected to require a smaller sample size.

Data on a range of life domain satisfactions in England (Defra, 2007) found that satisfaction with one’s community ranked lowest of all domains suggesting an important and urgent need to explore why many people are dissatisfied with this aspect of their life, and how this could be addressed. There may be concern that intervention to enhance neighbourly ‘chat’ is not
something that should be the remit of government policy. However, the analysis in this chapter suggests that taking to neighbours can be influenced by house and street design, social harmony in the neighbourhood and district levels of contact with neighbours. These are all clearly areas within the remit of local and national government over which the individual is unlikely to have full control.

The use of subjective measures of well-being has potentially highlighted an area which has been given insufficient attention following the use of other well-being indicators. Findings from analysis of large-scale surveys require verification from other research techniques, in this case from experimental research projects. Whilst these results suggest it would be pre-emptive to make policy recommendations, as the body of evidence on the importance of contact with those living close by expands, recommendations for policy are likely to become more apparent.
### APPENDIX 7.A

Table 38: OLS-FE and IV-FE for life satisfaction, including talking to neighbours

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (se)</th>
<th>Model 2 (se)</th>
<th>Model 5 (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talk to neighbours: most days</td>
<td>0.0513 (.019)**</td>
<td></td>
<td>0.1742 (.104)</td>
</tr>
<tr>
<td>Talk to neighbours: weekly</td>
<td>0.0105 (.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talk to neighbours: monthly (ref)</td>
<td>-0.0841 (.025)**</td>
<td>0.3202 (.018)**</td>
<td></td>
</tr>
<tr>
<td>Talk to neighbours: never</td>
<td>-0.0487 (.050)</td>
<td>0.2123 (.015)**</td>
<td></td>
</tr>
<tr>
<td>Talking to the neighbours scale</td>
<td>0.0364 (.008)**</td>
<td>0.2122 (.014)**</td>
<td></td>
</tr>
<tr>
<td>Log of household income</td>
<td>0.0131 (.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Living comfortably’</td>
<td>0.3227 (.017)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Doing alright’ financially</td>
<td>0.2122 (.014)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Coping’ financially</td>
<td>-0.2938 (.031)**</td>
<td>-0.2931 (.031)**</td>
<td>-0.2844 (.035)**</td>
</tr>
<tr>
<td>‘Finding it quite difficult’ financially</td>
<td>-0.6846 (.058)**</td>
<td>-0.6849 (.058)**</td>
<td>-0.6935 (.062)**</td>
</tr>
<tr>
<td>Education: Below O level qualifications or none (ref)</td>
<td>0.0505 (.068)</td>
<td>0.0494 (.068)</td>
<td>0.0124 (.075)</td>
</tr>
<tr>
<td>Education: O levels or similar</td>
<td>0.0879 (.073)</td>
<td>0.0870 (.073)</td>
<td>0.0224 (.081)</td>
</tr>
<tr>
<td>Education: Degree or similar</td>
<td>0.1363 (.062)*</td>
<td>0.1358 (.062)*</td>
<td>0.1130 (.069)</td>
</tr>
<tr>
<td>High hours carer (&gt; 50 hrs/week)</td>
<td>-0.2421 (.059)**</td>
<td>-0.2422 (.059)**</td>
<td>-0.2886 (.064)**</td>
</tr>
<tr>
<td>Marriage status: Married (ref)</td>
<td>-0.2497 (.047)**</td>
<td>-0.2502 (.047)**</td>
<td>-0.2333 (.052)**</td>
</tr>
<tr>
<td>Marriage status: Divorce/Sep.</td>
<td>0.0153 (.029)</td>
<td>0.0152 (.029)</td>
<td>0.0117 (.033)</td>
</tr>
<tr>
<td>Marriage status: Cohabiting</td>
<td>-0.3493 (.066)**</td>
<td>-0.3505 (.066)**</td>
<td>-0.4047 (.070)**</td>
</tr>
<tr>
<td>Marriage status: Widowed</td>
<td>-0.1189 (.047)*</td>
<td>-0.1200 (.047)*</td>
<td>-0.1352 (.055)**</td>
</tr>
<tr>
<td>Marriage status: Never married</td>
<td>-0.1955 (.028)**</td>
<td>-0.1956 (.028)**</td>
<td>-0.1800 (.029)**</td>
</tr>
<tr>
<td>Health: Problems walking</td>
<td>-0.0467 (.020)*</td>
<td>-0.0466 (.020)*</td>
<td>-0.0435 (.021)*</td>
</tr>
<tr>
<td>Health: Hospital stay</td>
<td>-0.1208 (.031)**</td>
<td>-0.1207 (.031)**</td>
<td>-0.0928 (.032)**</td>
</tr>
<tr>
<td>Health: Disabled</td>
<td>-0.4778 (.029)**</td>
<td>-0.4781 (.029)**</td>
<td>-0.4656 (.031)**</td>
</tr>
<tr>
<td>Health: SRH very poor or poor</td>
<td>-0.1931 (.015)**</td>
<td>-0.1931 (.015)**</td>
<td>-0.1858 (.016)**</td>
</tr>
<tr>
<td>Health: SRH fair</td>
<td>0.1045 (.013)**</td>
<td>0.1045 (.013)**</td>
<td>0.0986 (.014)**</td>
</tr>
<tr>
<td>Health: SRH excellent</td>
<td>-0.0236 (.016)</td>
<td>-0.0236 (.015)</td>
<td>-0.0298 (.018)</td>
</tr>
<tr>
<td>Number children: 12-18 years</td>
<td>-0.0154 (.014)</td>
<td>-0.0158 (.014)</td>
<td>-0.0191 (.017)</td>
</tr>
<tr>
<td>Number children: 5-11 years</td>
<td>-0.0118 (.016)</td>
<td>-0.0124 (.016)</td>
<td>-0.0259 (.019)</td>
</tr>
<tr>
<td>Number of adults</td>
<td>-0.0958 (.044)*</td>
<td>-0.0963 (.044)*</td>
<td>-0.1342 (.050)**</td>
</tr>
<tr>
<td>Number of adults squared</td>
<td>0.0555 (.007)</td>
<td>0.0555 (.007)</td>
<td>0.0128 (.008)</td>
</tr>
<tr>
<td>Sees friends/family most days</td>
<td>0.0581 (.018)**</td>
<td>0.0578 (.018)**</td>
<td>0.0273 (.028)**</td>
</tr>
<tr>
<td>Sees friends/family weekly</td>
<td>0.0235 (.016)</td>
<td>0.0231 (.016)</td>
<td>0.0116 (.019)</td>
</tr>
<tr>
<td>Sees friends/family monthly (ref)</td>
<td>-0.0498 (.034)</td>
<td>-0.0501 (.034)</td>
<td>-0.0348 (.037)</td>
</tr>
<tr>
<td>Sees friends/family &lt; monthly</td>
<td>-0.0447 (.182)</td>
<td>-0.0447 (.182)</td>
<td>-0.0076 (.203)</td>
</tr>
<tr>
<td>Employment: Employed (ref)</td>
<td>-0.2099 (.058)**</td>
<td>-0.2097 (.058)**</td>
<td>-0.2141 (.061)**</td>
</tr>
<tr>
<td>Employment: Long term sick</td>
<td>0.1079 (.033)**</td>
<td>0.1076 (.033)**</td>
<td>0.1058 (.037)**</td>
</tr>
<tr>
<td>Employment: Retired</td>
<td>-0.0925 (.047)*</td>
<td>-0.0924 (.047)*</td>
<td>-0.1034 (.052)*</td>
</tr>
<tr>
<td>Employment: Unemployed</td>
<td>0.2249 (.052)</td>
<td>0.2254 (.052)**</td>
<td>0.2173 (.054)**</td>
</tr>
<tr>
<td>Employment: Maternity leave</td>
<td>0.0253 (.029)</td>
<td>0.0249 (.029)</td>
<td>0.0171 (.031)</td>
</tr>
<tr>
<td>Employment: Self-employed</td>
<td>-0.0347 (.032)</td>
<td>-0.0349 (.032)</td>
<td>-0.0237 (.035)</td>
</tr>
<tr>
<td>Employment: Family carer</td>
<td>0.0272 (.087)</td>
<td>0.0268 (.087)</td>
<td>0.0180 (.089)</td>
</tr>
<tr>
<td>Employment: Other activity</td>
<td>0.0022 (.019)</td>
<td>0.0026 (.019)</td>
<td>0.0181 (.021)</td>
</tr>
<tr>
<td>Noisy neighbours</td>
<td>-0.0034 (.016)</td>
<td>-0.0035 (.016)</td>
<td>-0.0002 (.018)</td>
</tr>
<tr>
<td>Noisy street</td>
<td>-0.0138 (.015)</td>
<td>-0.0140 (.015)</td>
<td>-0.0189 (.016)</td>
</tr>
</tbody>
</table>
Table 39: OLS-FE and GLS-RE on 'talking to neighbours'

<table>
<thead>
<tr>
<th>Model 4: FE (se)</th>
<th>Model 5: RE (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life satisfaction last year (1-7 scale)</td>
<td>0.0138 (0.004)**</td>
</tr>
<tr>
<td>Log of household income</td>
<td>-0.0117 (0.009)</td>
</tr>
<tr>
<td>Education: O levels or similar</td>
<td>0.0489 (0.063)</td>
</tr>
<tr>
<td>Education: A levels or similar</td>
<td>0.0763 (0.068)</td>
</tr>
<tr>
<td>Education: Degree or similar</td>
<td>0.0378 (0.058)</td>
</tr>
<tr>
<td>High hours carer (&gt; 50hrs/week)</td>
<td>0.0197 (0.035)</td>
</tr>
<tr>
<td>Marriage status: Divorce/Separated</td>
<td>-0.0619 (0.034)</td>
</tr>
<tr>
<td>Marriage status: Cohabiting</td>
<td>-0.0557 (0.024)*</td>
</tr>
<tr>
<td>Marriage status: Widowed</td>
<td>0.0383 (0.039)</td>
</tr>
<tr>
<td>Marriage status: Never married</td>
<td>-0.0595 (0.041)</td>
</tr>
<tr>
<td>Health: Problems walking</td>
<td>-0.0245 (0.017)</td>
</tr>
<tr>
<td>Health: Hospital stay</td>
<td>0.0117 (0.013)</td>
</tr>
<tr>
<td>Health: Disabled</td>
<td>-0.0519 (0.020)**</td>
</tr>
<tr>
<td>Health: SRH very poor or poor</td>
<td>-0.0437 (0.018)*</td>
</tr>
<tr>
<td>Health: SRH fair</td>
<td>-0.0239 (0.010)*</td>
</tr>
<tr>
<td>Health: SRH excellent</td>
<td>-0.0120 (0.010)</td>
</tr>
<tr>
<td>Age 21 to 25</td>
<td>0.1914 (0.091)*</td>
</tr>
<tr>
<td>Age 26 to 30</td>
<td>0.2800 (0.096)**</td>
</tr>
<tr>
<td>Age 31 to 35</td>
<td>0.3251 (0.100)**</td>
</tr>
<tr>
<td>Age 36 to 40</td>
<td>0.3360 (0.104)**</td>
</tr>
<tr>
<td>Age 41 to 45</td>
<td>0.2970 (0.108)**</td>
</tr>
<tr>
<td>Age 46 to 50</td>
<td>0.2926 (0.113)**</td>
</tr>
<tr>
<td>Age 51 to 55</td>
<td>0.3121 (0.119)**</td>
</tr>
<tr>
<td>Age 56 to 60</td>
<td>0.3090 (0.124)*</td>
</tr>
</tbody>
</table>

88 Hansen J test is a test for overidentifying restrictions and is consistent in the presence of heteroskedasticity, asymptotically $\chi^2$ distributed, under the null hypothesis of validity. Hence the instruments are uncorrelated with the error term and correctly excluded from the initial model.

89 This is an $F$ test of exogeneity of the regressors in the IV-FE model where the null hypothesis is that the OLS-FE estimators are consistent (Davidson and MacKinnon, 1993).
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 61 to 65</td>
<td>0.3682</td>
<td>(.131)**</td>
</tr>
<tr>
<td>Age 66 to 70</td>
<td>0.4144</td>
<td>(.139)**</td>
</tr>
<tr>
<td>Age 71 to 75</td>
<td>0.4131</td>
<td>(.146)**</td>
</tr>
<tr>
<td>Age 76 to 80</td>
<td>0.4035</td>
<td>(.154)**</td>
</tr>
<tr>
<td>Age 80+</td>
<td>0.2927</td>
<td>(.163)*</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number children, 12-18</td>
<td>0.0150</td>
<td>(.012)</td>
</tr>
<tr>
<td>Number children, 5-11</td>
<td>0.0504</td>
<td>(.012)**</td>
</tr>
<tr>
<td>Number children, 0-4</td>
<td>0.0626</td>
<td>(.013)**</td>
</tr>
<tr>
<td>Adults</td>
<td>0.1137</td>
<td>(.034)**</td>
</tr>
<tr>
<td>Adults squared</td>
<td>-0.0193</td>
<td>(.006)**</td>
</tr>
<tr>
<td>Employment status: Long term sick</td>
<td>0.0470</td>
<td>(.035)</td>
</tr>
<tr>
<td>Employment status: Retired</td>
<td>0.0957</td>
<td>(.023)**</td>
</tr>
<tr>
<td>Employment status: Unemployed</td>
<td>0.0906</td>
<td>(.032)**</td>
</tr>
<tr>
<td>Employment status: Maternity leave</td>
<td>0.0627</td>
<td>(.046)</td>
</tr>
<tr>
<td>Employment status: Self-employed</td>
<td>0.0422</td>
<td>(.024)</td>
</tr>
<tr>
<td>Employment status: Family carer</td>
<td>0.0530</td>
<td>(.022)**</td>
</tr>
<tr>
<td>Employment status: Other activity</td>
<td>0.0866</td>
<td>(.058)</td>
</tr>
<tr>
<td>Noisy neighbours</td>
<td>-0.0536</td>
<td>(.015)**</td>
</tr>
<tr>
<td>Noisy street</td>
<td>-0.0448</td>
<td>(.012)**</td>
</tr>
<tr>
<td>Crime and vandalism</td>
<td>0.0090</td>
<td>(.011)</td>
</tr>
<tr>
<td>Years at address</td>
<td>-0.00003</td>
<td>(.000)**</td>
</tr>
<tr>
<td>Time working</td>
<td>-0.0007</td>
<td>(.001)</td>
</tr>
<tr>
<td>Time working squared</td>
<td>-0.0005</td>
<td>(.001)</td>
</tr>
<tr>
<td>House type: Semi-detached</td>
<td>0.0695</td>
<td>(.021)**</td>
</tr>
<tr>
<td>House type: Terraced</td>
<td>0.0838</td>
<td>(.025)**</td>
</tr>
<tr>
<td>House type: Flat</td>
<td>0.0412</td>
<td>(.034)</td>
</tr>
<tr>
<td>House type: Other</td>
<td>0.0581</td>
<td>(.045)</td>
</tr>
<tr>
<td>Tenure: Rented public</td>
<td>0.1175</td>
<td>(.036)**</td>
</tr>
<tr>
<td>Tenure: Rented private sector</td>
<td>-0.1249</td>
<td>(.035)**</td>
</tr>
<tr>
<td>House shares a garden</td>
<td>0.0533</td>
<td>(.023)*</td>
</tr>
<tr>
<td>District av. talking to neighbours scale</td>
<td>0.1971</td>
<td>(.021)**</td>
</tr>
<tr>
<td>Urban</td>
<td>-0.4527</td>
<td>(.182)*</td>
</tr>
<tr>
<td>Wave dummies</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Regional dummies</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R squared</td>
<td>Within: 0.0226</td>
<td>Overall: 0.0944</td>
</tr>
<tr>
<td></td>
<td>F(79,36800) = 8.92</td>
<td>Wald chi2(80) = 2140</td>
</tr>
<tr>
<td>Observations</td>
<td>46402</td>
<td>46402</td>
</tr>
<tr>
<td>Groups</td>
<td>9523</td>
<td>9523</td>
</tr>
</tbody>
</table>

Note: Waves 7-11&13-15. Wave 12 is not included as there is no lag of life satisfaction in this wave.
Chapter 8: SWB as a metric of value

8.1 Introduction

This chapter looks at valuing benefits associated with non-market goods, which has traditionally involved the use of willingness to pay (WTP) or, in the case of health care, public preferences towards different states. If SWB measures offer an alternative means of valuing non-market goods this has immense importance to policy makers. Chapter One noted that government’s need to make comparisons across a broad range of possible interventions and this ideally requires benefits to be expressed as a single metric. The contingent valuation (CV) method discussed in Chapter One is widely used. For example, in the UK the Environmental Act of 1995 imposed duties upon the Environmental Agency to account for the costs and benefits arising from its policies (Bateman et al., 2000). Stated and revealed reference techniques have “enjoyed an unprecedented increased in application both in the developed and developing world” (Bateman et al., 2000: 1).

WTP studies have faced many criticisms, some of which have been somewhat addressed by improvements in study design (Carson et al., 2001; Arrow et al., 1993; Carson, 1997; Bateman et al., 2002), others which remain unsolved.

WTP studies struggle to deal with the fact that preferences may be artificially created during the survey and may not represent true preferences. Firstly, individuals may adopt strategic behaviour (Carson, 1997), answering higher or lower than their true WTP based on an attempt to influence study outcomes. Secondly, valuations given may be based on ‘warm glow’ feelings (Andreoni, 1989), or moral obligation (Diamond and Hausman, 1994).

WTP studies have been shown to suffer from insensitivity to scope, whereby the individual’s WTP is unrelated to the size of the benefit being valued (Kahneman and Ketsch, 1992). This has been found to be particularly problematic when individuals are asked to value small changes in the risk of negative events such as fatal traffic accidents or being a victim of violent crime (Loomes, 2006). Insensitivity to scope includes ‘embedding effects’ which occur when the respondent finds it difficult to isolate a specific case from overall considerations. For example, a respondent may react to a good’s general “symbolic meaning instead of to the specific levels of provision described” (Mitchell and Carson, 1989: 249).

Responses may be subject to focusing effects whereby the question concentrates the individual’s attention upon the good being valued. Focusing effects concentrate a respondents’
attention on how valuable the non-market good is, and not on other attributes of importance in their life, and nothing in life is as important as when you are thinking about it (Schkade and Kahneman, 1998).

A further difficulty that has yet to be resolved is the reliance in WTP studies on individuals’ prediction of hypothetical states. It is increasingly recognized that people have difficulties accurately predicting how they will feel under future scenarios (Kahneman and Sugden, 2005). We fail to predict adaptation to new circumstances, being influenced unduly by our immediate affective reaction to choices we face (Wilson and Gilbert, 2003). We overestimate our own willpower (Ausubel, 1991), and we fail to predict how the value of items changes once they become our possession (Kahneman et al., 1991). Our choices are influenced by our faulty recall and memories of previous experiences (Kahneman, 1999) and by our expectations rather than by actual occurrences (Mitchell et al., 1997). WTP studies are addressing this through the provision of additional information describing the non-market good, yet this is still unlikely to give a convincing picture of how the individual will feel under different future circumstances.

Even if CV studies could improve the accuracy of peoples’ predictions of how they would feel under future scenarios, if changes in income are a poor measure of changes in well-being, and increasingly so at high levels of income, CV studies, no matter how well conducted, will not provide a measure of individual change in well-being. This thesis has highlighted various deficiencies with the use of income as a measure of well-being. Chapter Three showed that income is out of line with a range of other well-being measures, suggesting a lack of convergent validity. Chapter Five argued that income failed to meet a set of criteria for a useful measure of well-being for public policy. Chapter Six discussed evidence that showed the importance of past incomes and the income of others to SWB suggesting that income can not be treated independently as a measure of well-being. If change in income does not reflect a change in well-being, WTP may measure how much, at a moment in time, an individual thinks they want a particular good or service, in terms of their hypothetical monetary sacrifice, but it cannot be a measure of the additional well-being from that good or service.

This understanding is not new. Pollack (1976) for example, modelled utility as function of both current consumption and a weighted aggregated of lagged consumption of self and others. From the interdependence of preferences, he concluded:

“…we cannot base judgements about individual welfare on the preference ordering revealed by market behaviour; to make judgements about welfare of an individual we need more information” (Pollack, 1976: 320).

However, the case is now made with empirical support based large datasets leaving no doubt to the accuracy of Pollack’s claim.
Furthermore, evidence supporting a relative income effect shows that individual WTP cannot be aggregated into average WTP since some of the (small) well-being impact from income is likely to arise from changes in relative income. As noted by Frank and Sunstein the numbers generated for WTP for safety measures (which would be implemented socially not just to the individual) are too low “insofar as they fail to take account of concerns about relative economic position, and that a serious conceptual defect is thus inherent in current approaches” (Frank and Sunstein, 2001: 326).

It is therefore clear that WTP methods have severe fundamental weaknesses. Consequently policy makers should look to alternative means for valuing non-market goods. Two alternative options for valuating non-market goods present themselves from the SWB literature. The first is the direct change in SWB associated with a change in the non-market good, and the second is the monetary equivalent of this change. This chapter will explore these two alternatives and consider the limitations and the advantages of each.

It is important to take stock of this new methodology for valuing non-market goods, within the context of a clear concept of well-being that this methodology is founded upon. Data showing both the responsiveness of SWB measures to non-market goods, and providing monetary equivalents, is becoming increasingly available. Where monetary figures are provided (particularly if they appear reasonable) these will be very attractive for policy makers. It is therefore essential that that there is clarity on exactly what such numbers are telling us.

### 8.2 SWB as a direct metric of value

The change in SWB following a change in a variable of interest can be estimated either from analysis of large datasets (preferably panel, or if not, datasets with detailed personality variables), or from experimental studies.

Comparisons between changes in well-being from different measures require standardisation and they would need to be presented within the context of both the scale of the instrument and its usual distribution, with outcomes shown in terms of ‘effect size’. Synthesis of results from different studies, and cross-validation with studies using different methodologies, should generate confidence in the effect size. In general, a move towards greater consistency across the SWB measures will aid comparison and increase familiarity with SWB as a measure of value.

One major advantage of taking well-being change from this kind of dataset is that respondents do not need to be made aware of the good that is being valued at the time of the survey and where necessary the focusing effect can be diverted by asking a range of other questions (as done by van Praag and Baarsma, 2005).
Whilst avoiding many of the pitfalls of CV studies, the use of a direct measure of SWB change faces seven additional problems. Firstly, appropriate data may not exist to enable an assessment of SWB change; secondly, the individual effect on SWB may be too small to be detected; thirdly, the benefits to be valued may not be part of prudential value; fourthly, the benefits of the non-market good may arise, in part, at some time in the future; fifthly, the non-market good may, in part, be already compensated for within market transactions; sixthly, the lack of understanding of the determinants of well-being and direction of causality risks the coefficients in the SWB regression being subject to bias; and lastly, if we are able to detect a change in SWB there is no consensus on how much a well-being change should be worth. These problems shall be considered in turn.

i) Appropriate data may not exist

The SWB method will be limited to valuing attributes that already exist for at least some people and where there is variation in the non-market good either across individuals or over time. If we wish to value a non-market good, such as a new piece of infrastructure, it would be necessary to infer values found from areas where similar infrastructure was present. Valuations derived in other contexts, particularly internationally, may be problematic since the impact of economic, social and environmental characteristics is likely to vary depending on context. Analysis of existing data may not be representative of future change, particularly if the implementation has yet to occur within a comparable location or subgroup. Although benefits transfer is a well-being known problem facing CV studies (Bateman et al., 2000), it is particularly pertinent to SWB given the size of questionnaire required to ensure necessary control variables, and the sample size required in order to obtain significant coefficients.

ii) The individual effect may be too small to detect

The impact of policy that is expected to have a small individual impact on well-being may be difficult to identify, even if it impacts upon a large number of people. Potentially the SWB method can be applied to any non-market good that is thought to contribute towards well-being. However, if the dependent variable is a single-item life satisfaction question, we may not expect attributes that have a small impact on SWB to be picked up. Sensitivity to important changes in well-being can be potentially increased through increasing the number of response outcomes to single-item questions (Cummins, 2003), or through the use of multi-item scales, such as the Satisfaction With Life Scale (Diener et al., 1985) or domain-weighted satisfactions. Furthermore, large datasets enable researchers to identify small, yet statistically significant,

---

90 This is a good example of the potential distinction between economic or political significance and statistical significance (see McCloskey and Siegler, forthcoming).
improvements in SWB. Nevertheless, for a non-market good in which the typical WTP valuation is a few pounds, identifying this valuation through changes in life satisfaction responses would seem unlikely. Where a particular good is to be valued, and that good is not treated as a usual control variable within SWB models, its inclusion within existing large datasets is unlikely, suggesting the need for additional data collection, thereby generating a constraint on sample size.

Smaller changes in SWB can be detected using moment-to-moment measures. Experience sampling methods (ESM) and the day reconstruction method (DRM) (discussed in Chapter Two) provide promising ways of doing this. However, ESM and DRM, as they are currently used, tap into how people feel and not how they think about their lives; that is, they reflect a hedonic rather than an evaluative account of SWB. If well-being is conceived as an individual’s evaluation of their life, ESM and DRM in their current form, represent incomplete picture of well-being and would need to be extended to capture the more reflective parts of SWB.

Even with the most sensitive of measurement tools, a direct change in individual SWB cannot be used to value goods that do not impact on individual SWB. This is likely to be particularly problematic for environmental CBA, such as protecting a species from extinction.

iii) Benefits may not be part of prudential value

Krutilla (1967) argues that many people value natural wonders simply for their existence, aside from any personal benefit they may receive from the good. Environmental economists have conducted a number of studies attempting to value the existence or non-use value of environmental goods (Carson et al., 1999). Pure existence value may show up in an individual’s SWB when she thinks about the consequent benefits that result from the preservation of a rainforest, a species of dolphin etc. It may also be the case that current SWB is influenced by living in the kind of society which chooses to respect the environment and which cares about the well-being of future generations.

It is also likely that people have preferences towards preventing species extinction that go beyond their expectation of the impact that holding these values would have on their own well-being. Because CV is founded on a preference satisfaction approach, in which people can hold preferences towards things other than their own well-being, extending beyond prudential value, a preference satisfaction approach can be used to value such attributes (although as discussed in Chapter Five, the inclusion of non-prudential values results in a measure of well-being which is less suited for public policy needs). Since the direct SWB method is based on an individual’s assessment of their prudential value (how well life is going for them), it does not necessarily tap
into preferences people hold towards things other than their own utility. Kahneman and Sugden (2005) note the limitation of using moment-based happiness measures to assess pure existence value. Individuals may wish for resources to be allocated to a proposed intervention not because it impacts upon their well-being, but because of its relevance to other values that they hold.

iv) Benefits may arise some time in the future

If a non-market good has benefits for future people these would need to be explicitly considered. The SWB methodology cannot inform policy as to the weight which should be given to the well-being of future people. However, the SWB methodology could be complemented by finding out the extent to which individuals are willing to trade-off their current well-being for the future well-being of other people. Similarly, to value an attribute which has future well-being implications requires an additional step of considering the pattern of future benefits, and the rate at which it is appropriate to trade-off an individual’s current benefit with their future benefit. Of course, governments may adopt a normative stance in which future well-being is equally valued to current well-being regardless of any preferences individuals hold towards either discounting their own future well-being or the well-being of future members of society. In terms of increased transparency, there is much appeal in disaggregating benefits of a non-market good into benefits experienced by current population, experienced in the future by the current population and experienced in the future by future members of the population.

v) The benefit may be partly accounted for within market transactions

A further problem arises in valuing a non-market good which may be partly attained via the market. Consider the example of street noise, which can be thought of as a classic externality, therefore something in which policy makers are likely to show interest. Peace and quiet may be indirectly purchased via the housing market (Day et al., 2003). Indeed in perfect markets we would expect an equilibrium housing market to fully account for all factors relevant in the choice of where to live. Living on a noisy street may be compensated for by cheaper rent, a larger house etc. Inflexibilities and imperfections in the housing market imply that the impact of noise is unlikely to be fully accounted for within rental and house prices. Consequently, there may be a residual noise effect that cannot be purchased in the market.

Where change in SWB is estimated from the coefficient on street noise in a typical SWB regression, in which income, housing costs and housing quality are controlled for this would give a more complete picture of the SWB loss caused by street noise, yet not all of that loss can be treated as a non-market good. However, there may be other compensations for street noise.
(which may or may not be purchased via the market), such as lower commuting time, which are not controlled for. Consequently, this missing variable would lead to an underestimate of the full impact of street noise. Understanding the full extent of the cost imposed by street noise could use a combination of SWB valuation and hedonic pricing in order to capture the full cost of such a non-market good, and considerable care will be needed in ensuring the appropriate choice of controls. However, this re-introduces the problems with WTP discussed above. Another alternative may be to consider the impact of street noise only upon those who are unable to move house.

**vi) Coefficients in SWB models may be bias**

Research is still continuing to identify the factors that have a significant relationship with SWB, such as community relationships (Dolan *et al.*, 2006a), studies that do not control for such factors where they may be correlated to the variable of interest risk bias coefficients. Furthermore, there is little conclusive evidence on the direction of causality of many variables. For example, a significant coefficient on street noise is still only evidence of association rather than causality. As discussed in the previous chapter, there is not always a statistical quick fix for this, although increased use of instrumental variable techniques may go some way to exploring cases where coefficients may be bias. Estimates from experimental or quasi-experimental trials are needed to complement SWB valuations from regression analysis using large household surveys.

**vii) Are the benefits worth the costs?**

The last concern relates to the need to know whether an intervention is worth its costs. Although it is possible to compare across alternatives using the incremental cost per unit increase in SWB, the valuation metric of SWB offers little knowledge of how much a SWB change should be worth, or society’s WTP for a unit increase in SWB. This then appears to provide less information than more conventional CV studies. If we require an answer as to whether benefits exceed costs, this requires a monetary expression of benefits.

### 8.3 Indirect valuation via income compensations

The lack of a clear value for changes in SWB can be overcome if all benefits are expressed in monetary terms enabling benefits to be directly compared to costs.

---

91 SWB valuations therefore suffer from the opposite problem to hedonic pricing which assumes that the total effect of an attribute can be identified indirectly via the market.
Recently economists have begun valuing non-market goods by considering the effect on an individual’s SWB of income and the non-market good, and then estimating the required income compensation that would hold SWB constant following a change in the non-market good (Clark and Oswald, 2002). This method has been used to give a monetary valuation for a range of non-market goods, including marriage (Clark and Oswald, 2002), informal care (van den Berg and Ferrer-i-Carbonell, forthcoming), aircraft noise (van Praag and Baarsma, 2005), air pollution (Welsch, 2002; 2006), the psychological cost of unemployment (Winkelmann and Winkelmann, 1998), chronic diseases (Ferrer-i-Carbonell and van Praag, 2001), cardiovascular disease (Groot et al., 2004), friendships (Powdthavee, 2007), trust in the workplace (Helliwell and Huang, 2006), death of a friend or family member (Oswald and Powdthavee, forthcoming) and terrorist activities (Frey et al., 2004).

The method assumes that well-being can be represented as a function of income (Y), the non-market good of interest (X) and a vector of other influential social, economic and environmental factors (Z)

\[
SWB = \beta_1 X + \beta_2 Y + \beta_3 Z'
\] (1)

As with contingent valuation, there are two perspectives from which to assess the value of the non-market good. Following a change in the non-market good (X) and a new SWB level, equivalent variation estimates the change in income which would also have taken the individual to the new SWB level, i.e.

\[
SWB_1(X + \Delta X + Y + Z') = SWB_1(X + \Delta Y + Y + Z')
\] (2)

The compensating variation estimates the change in income necessary to hold SWB constant following a change in the non-market good i.e.

\[
SWB_0(X + Y + Z') = SWB_0(X + \Delta X + Y + \Delta Y + Z')
\] (3)

The ‘shadow price’ of the non-market good which represents both the equivalent variation and compensating variation can be found when \(\Delta X\) tends to zero, or from the slope of the indifference curve, i.e.

\[
\frac{\delta SWB}{\delta X} = \frac{\beta_1}{\beta_2}
\]
This method therefore relies upon the validity of estimates of the impact of both the non-market good and income upon SWB. Consequently is subject to the problems facing the use of direct change in SWB, plus further problems arising from need for an accurate measure of the impact of income upon SWB.

As discussed in Chapter Two income may be an inaccurate measure of availability of resources. Income is notoriously difficult to measure accurately and many surveys use bands for income (Frey et al., 2004). Small scale surveys focused on particular non-market goods may feel constrained in asking detailed income questions, due to time and the need for high response rates. Moreover, individual income may not represent command over resources: household income, equivalised household income, or some measure of wealth or permanent income may do so more accurately. For example, Headey and Wooden (2004) found that wealth was slightly more important than income for life satisfaction and income was no longer significant when using mental health as a measure of well-being. One solution is to restrict the age of the sample to those for whom current income is more likely to be an adequate reflection of consumption (such as to 30-55 year olds as done by Layard et al., 2007). However, this is problematic if the non-market good benefits individuals outside that age group. Furthermore, if this age group show a different relationship between income and SWB without further research it will be unclear if this is due to inaccuracies in income as a proxy for consumption in other age groups or reduced importance of income. It is also not clear what approach should be taken if household and individual income show a different rate of substitution to the non-market good.

The problem mentioned above relating to a potential lack of certainty in the determinants of SWB or direction of causality is particularly acute in relation to income. The coefficient on income is often not robust to the inclusion of other variables. There is also ongoing uncertainty surrounding the direction of causality of income and such endogeniety may over or underestimate the income coefficient.

A number of studies have found that the coefficient on income changes considerably once additional controls have been added. The exclusion of variables to which income is correlated, such as health and individual personality, generate bias in the income coefficient, consequently, it shows considerable change following their inclusion. For example, using Hungarian household survey data, Lelkes (2006) finds that the coefficient on log of equivalised household income (from an ordered logit on life satisfaction) falls from 0.824 to 0.626 once employment status, health and housing quality are included. The change in magnitude of the income coefficient is particularly strong when country and individual fixed effects are controlled for. Using data from 12 European nations, Di Tella et al. (2003) find that the coefficient on per capita GDP changes from 1.408 to 1.132 when country specific time effects are included, and
then to 1.020 once lagged GDP/capita is also included. Similarly, Graham and Pettinato (2001) find the coefficient on log of wealth falls from 0.463 to 0.264 after including country dummies. Using German data, Ferrer-i-Carbonell and Frijters (2004) find the coefficient on log household income falls from 0.38 to 0.11 once fixed effects are accounted for. Luttmer (2005) finds similar results.

If there is a desire to express benefits in monetary terms, then we need to estimate a robust income coefficient by determining the most appropriate utility function and establishing what other regressors should be included. Neither of these issues has been resolved, either in theory or in practice. However, we do know that studies which do not fully account for individual or country fixed effects (Ferrer-i-Carbonell and Frijters, 2004), and which do not include variables that are now known to affect SWB, are unlikely to generate a reliable monetary valuation of the non-market good.

Whilst we do not want the income coefficient to be picking up the effect of non-income factors, we do want the income coefficient to capture all direct and indirect effects of income. Controls such as house type, tenure, and even health, are likely to capture an indirect income effect. Additional income leads to improved life circumstances across a range of attributes, particularly housing conditions. If the benefits of higher income (such as better health, living in a safer neighbourhood, owning your own home etc.) are controlled for through the addition of health status, housing tenure etc., then the coefficient on income will be an underestimate. A simple SWB model identifies the impact of income upon current levels of SWB, yet many of the benefits of higher income may arise in future periods, e.g. through improved health. Ideally all indirect effects of income should be accounted for, however, this would still inadequately incorporate future benefits arising from current income.\(^\text{92}\)

There is a strong theoretical basis for the inclusion of relative income within the utility function and many empirical studies have shown that relative income matters in addition to current income (see Chapter Six).

An income coefficient that incorporated both the absolute and relative effect of income is theoretically closest to WTP or WTA since from the individual’s perspective payment or compensation would alter both their relative and absolute position.

However, when aggregating across a group of people, it makes little sense to value the non-market good based on a zero sum item. This highlights the inadequacy of income change as a

\(^{92}\) This framework also only identifies benefits in terms of SWB, yet income may bring other (non-prudential) benefits which individuals value (Kimball and Willis, 2006).
unit of value. An individual may be compensated by an increase in rank position within the income distribution, however, this change is not a useful metric with which to compare the social value of a non-market good. If we wish to aggregate the compensation required for the non-market good this should be based only on the impact of absolute income.

Controlling for relative income and the form in which it is included makes a substantial difference to the size of the income coefficient. Using pooled data from the US GSS, Blanchflower and Oswald (2004) find that the coefficient on log of household income per capita (from an ordered probit on how happy things are these days) falls from 0.2209 to 0.1723 once relative income (household income per capita/state income per capita) is included and further to 0.1349 once regional house price index is also included. Similarly, Ferrer-i-Carbonell (2005) using panel data from Germany, finds that the coefficient on log of family income (from an ordered probit on present happiness) remains at 0.248 once log of average income of a reference group (defined by similar education, age and region) is included, but falls to 0.109 (and is no longer significant) when the difference between family income and reference income is included. Indeed, as seen in Chapter Six, in the BHPS data once the ratio of household to reference income is controlled for own household income has a negative coefficient.

Given that controlling for reference income has a considerable impact, it will be necessary to be clear about whether the income effect required in the compensating equation is the absolute income effect only, or an income effect which includes changes in rank position. If compensation is actually to be paid, particularly if the compensation is derived for people living close to each other, and hence deemed to be within each others reference group (as the compensation for airport noise suggested by van Praag and Baarsma (2005) would be), it is important to account for relative income effects. How relative income is modelled, and who the appropriate reference groups should be, remains undetermined. However, income compensation studies need to be explicit about their treatment of relative income effects, and the normative implications of this.

Income compensations are generally estimated from static models, which should represent a steady state (for example, although there maybe adaptation to marriage with the benefits declining over time, the coefficient in the steady state model should represent the average benefit from being married). There may be adaptation to income (as discussed in Chapter Six) and to the non-market good, the steady state rate of substitution should pick up the long run trade-off between these two goods. However, if income compensations are being used for the purpose of determining compensation that is to be paid, where this amount is considerable (e.g. Oswald and Powdthavee, forthcoming), particularly if it is be given in a lump sum rather than annually and is out of line with usual income changes, it would seem problematic not to account
for additional adaptation to this income increase. It seems problematic to conceive of an individual being compensated in period 1, when that compensation will lower their SWB in period 2. Indeed, the whole concept of single period compensating income variation or equivalent income variation in either CV or SWB is undermined by a reference dependent relationship between income and utility (Munro and Sugden, 2003).

The relationship between income and SWB may be endogenous. It is possible, for example that the direction of causality runs from SWB to income (Graham et al., 2004; Lyubomirsky et al., 2005; Marks and Fleming, 1999; Schyns, 2001). There is longitudinal evidence to suggest that happier people earn more (Diener et al., 2002) and are less likely to lose their jobs (Verkley and Stolk, 1989). Alternatively, a third factor may be correlated with both income and SWB. A missing factor might be time invariant e.g. those with an extrovert personality may have both higher income (they may work better with people, for example) and have higher SWB (DeNeve and Cooper, 1998). If personality is not controlled for within the Z matrix or through the use of fixed effects, then the income coefficient will be biased. Alternatively, a missing factor may be time variant e.g. if an individual has a sick child, this may lower both their income and their SWB and, again, failure to control for this will result in biased estimates.

Oswald and Powdthavee (forthcoming) instrument for income in the BHPS using whether a pay slip is seen by the interviewee and the previous periods regional house prices. They find that instrumenting dramatically changes the size of the coefficient on the log of personal income from 0.091 (se 0.020) to 0.698 (se 0.097) in an OLS on life satisfaction and from 0.018 (0.017) to -0.818 (se 0.144) in an OLS-FE on the GHQ. They suggest this may be explained in terms of happy people working less to earn money. An underestimate of the income coefficient may also arise due to measurement error. Where this is random it is likely to bias the income coefficient towards zero. Alternatively, a positive measurement error on income may be more likely for those with relatively lower SWB reports (i.e. people who are feeling less positive about their lives may have a tendency to overestimate their income), which would lead to an underestimate of the income coefficient.

Luttmer (2005) finds that the coefficient on log household income rises from 0.121 (se 0.020) to 0.365 (se 0.100) once other income proxies (home value and tenure status e.g. renting) are removed and income is instrumented using predicted household earnings based on industry and occupation of the respondent and spouse and national earnings information by industry and occupation, however, it is unclear whether this is being driven by correcting for endogeneity or the removal of housing variables. Using a similar method, Luechinger (2007) instruments for income in the GSOEP using a predicted value of income based on 5000 industry-occupation cells along with region and year and the job tenure of either the main earner or the secondary
earner. Instrumenting brings a three-fold increase to the income coefficient. These changes in magnitude of the income coefficient will have a substantial impact upon the compensating variation.

The evidence suggests that income has a greater effect on SWB at lower levels of income (Biswas-Diener and Diener, 2001; Cummins, 2000; Diener and Seligman, 2004; Fahey and Smyth, 2004). Diminishing returns to income, as discussed in Chapter One, usually lead to income being included in logarithmic form. Van Praag and Baarsma (2005) point out that this has the politically problematic implication that compensation for an attribute that is detrimental to happiness would be greater for those on higher incomes. Of course, valuation exercises do not need to be concerned with actual compensation, and can take the average value, as is done in CV studies.

A further problem arises when income is modelled in logarithmic form, in that there will be asymmetry between the additional income, from the mean income, required to compensate for the presence the non-market good, and the reduction which would leave life satisfaction unchanged following the change in the non-market good. This would be avoided were income to be included in a linear fashion; however, this would overlook the diminishing marginal utility of income.

The marginal utility of income has also been shown to vary by initial level of SWB (Boes and Winkelmann, 2004) and by age group and relationship status (Cummins et al., 2004) and religious beliefs (Lelkes, 2002). This raises an important normative issue of whether the value of a non-market good should be higher for individuals for whom SWB is less responsive to changes in income. For example, if the elderly and higher income groups have a smaller marginal utility of income, then when valuing a non-market good which is mostly used by these groups, would it be most appropriate to use the specific income coefficient for the sub-set for whom the non-market good is relevant, or a population level income coefficient? If the latter, then the compensation would not imply returning the individual to their initial level of well-being, but the former would imply non-market goods used by these subgroups to have relatively greater value than non-market goods used by other groups.

Similar problems will plague the use of income compensation in the derivation of family equivalence scales (Schwarze, 2003; Rojas, 2007; van Praag and Ferrer-i-Carbonell, 2004). The use of equivalence scales which fully compensate for the presence of an additional child raises further normative issues. Do we wish to compensate (in welfare payment decisions or analysis of poverty and income inequalities) for either life satisfaction or income satisfaction by the presence of children, which are arguably a choice variable and whose existence is unlikely to be
well-being neutral for parents. Furthermore, children may change attitudes and expectations of financial circumstances.

If the presence of a child enhanced satisfaction with income (which may arise if parents give less emphasis to financial success and reach contentment on lower absolute income – despite their extra financial need) this would result in a negative compensation. That this is at least possible suggests this not a fruitful route to go down.

Welsch (2002, 2003) argues that his estimates of the value of air pollution represent the trade-off a representative individual is ‘willing to make’ between prosperity and the environment. However, an individual may experience greater SWB from living with less air pollution, without having any desires towards clear air. More generally, SWB methods may value attributes (such as talking to neighbours, or breathing clean air) that individuals do not have clear desires towards. Frey et al. (2004) also describe the SWB compensating income valuation as a ‘willingness to pay’ but there is no willingness being expressed by these compensating income valuations. We may think that because of the trade-off against income individuals should be willing to pay this amount on average, but we cannot avoid the fact that individuals – even if presented with the evidence – may not be willing to sacrifice the SWB equivalent income.

8.4 Exploring income compensation issues using BHPS.

Some light may be shed on some of the problematic issues of using income compensation with SWB data through exploring an example of an externality for which data is present in the BHPS. The example used here is living on a noisy street. This is particularly pertinent within the UK context since problems with noisy neighbours and people being rowdy in public places are showing an increasing trend.93 Although we may think of a quiet neighbourhood as something that can be purchased via the market for houses or rented property, there are also reasons why this has properties of a non-market good. Information asymmetries with the housing market mean that the buyer has little experience of noise on the street, particularly at night. High transactions costs result in rigidities in the market, and for many who rent from the public sector or housing associations opportunities to reveal choice are limited.

As noted in Chapter Seven, since wave six of the BHPS respondents have been asked whether their accommodation suffers from a range of problems, including street noise (other than noise

---

93 Those reporting a problem with noisy neighbours or loud parties rose from 8% in 1997 to 11% in 2007, similarly, those reporting a problem with drunk or rowdy people in public places rose from 22% in 2002 to 26% in 2007 (Defra, 2007).
from neighbours). Despite the increasing trend noted above those reporting this problem fell from around 21% in wave 6 to 15% in wave 15. As would be expected a higher percentage report a problem in metropolitan areas (20% in wave 15) than in non-metropolitan areas (14% in wave 15).

The SWB-income compensation valuation method faces a difficulty when the non-market good may itself be picking up income effects. The data show that household income is lower for those reporting street noise, and those reporting street noise are more likely to live in accommodation rented from the public sector or housing associations. Yet it is important that the income coefficient captures all the effect of a change in income.

If the street noise variable picks up income effects, this would potentially both give an overestimate to the coefficient on street noise and underestimate the impact of income, combining to have a double impact on the compensation variation. If other social conditions in the area are not controlled for, then street noise may pick up a host of other problems, such as fear of crime, dirt, vandalism etc. An attempt is made to control for the socio-economic characteristics by additionally controlling for noise from neighbours, crime and vandalism in the area and environmental and pollution problems. This may weaken the income coefficient slightly, but will help ensure that the coefficient on street noise is singling out just that.

Life satisfaction is therefore modelled as a function of street noise, log of household income (gross, current) and a range of control variables (those used in Chapter Three, plus controls for noise from neighbours, crime and vandalism and environment and pollution problems) using OLS fixed effects. When using life satisfaction as a dependent variable, street noise, although negative, is not well defined consequently no compensating income can reasonably be estimated (Table 40, row 1).

However, when using domain-weighted satisfaction (described in Chapter Three) street noise becomes significant (Table 40, row 2). The income coefficient is fairly small. This may in part be due to outlying income values, which may be a poor reflection of consumption for those individuals, suggesting the need to remove outliers. To address potential inaccuracies and measurement error within household income outliers are removed using the method applied by Layard et al. (2007). Outliers are identified based on those for whom the data inaccurately predicts household income (those with the largest 1% of positive residuals and largest 1% of negative residuals from a prediction of logged household income). Following their exclusion the size of the income coefficient doubles (Table 40, row 3). Consequently, the compensation for a noisy street also increases by a similar magnitude. Where 10% of outlying cases are removed (as done by Layard et al., 2007) the coefficient increases by a further 25% (Table 40, row 4).
However, if we wish to apply the method across wide range of the population, this level of exclusion seems excessive.

If the SWB compensation method requires controlling for the individual heterogeneity via the use of fixed effects analysis with panel data then potential applications will be limited. Where data is collected specifically for the purpose of valuing a non-market good, single period cross-section data is more likely to be available. The use of single period cross-section data can be explored through wave 15 which contains 15 questions pertaining to the ‘big five personality traits’; openness to experience, neuroticism, agreeableness, extraversion, and conscientiousness (Goldberg, 1990).

Using only wave 15 the income coefficient increases slightly, as would be expected if positive personality traits led to both increased income and increased SWB (Table 40, row 5). The coefficient on street noise also increases four fold, suggesting between person variance shows a stronger relationship between street noise and SWB than within person variance. Once personality is controlled for, the income coefficient drops a little, although the coefficient on street noise remains very similar (Table 40, row 6).

### Table 40: Regressions on life satisfaction and domain-weighted satisfaction

<table>
<thead>
<tr>
<th>Dependent variable (waves)</th>
<th>Income coeff. (se)</th>
<th>Street noise coeff. (se)</th>
<th>R²</th>
<th>Obs.</th>
<th>Equiv. units of total income.</th>
<th>Compensating income/year, estimated from mean income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. OLS-FE Life satisfaction (7-15)</td>
<td>.0385 (.0082)**</td>
<td>-.0069 (.0125)</td>
<td>0.0408</td>
<td>87349</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. OLS-FE Domain-weighted, (7-15)</td>
<td>.0315 (.0049)**</td>
<td>-.0162 (.0072)*</td>
<td>0.0764</td>
<td>84807</td>
<td>0.514</td>
<td>£19,727</td>
</tr>
<tr>
<td>3. OLS-FE Domain-weighted, excl. 2% (7-15)</td>
<td>.0612 (.0065)**</td>
<td>-.0146 (.0073)*</td>
<td>0.0761</td>
<td>83129</td>
<td>0.238</td>
<td>£7,884</td>
</tr>
<tr>
<td>4. OLS-FE Domain-weighted, excl. 10% (7-15)</td>
<td>.0811 (.0087)**</td>
<td>-.0156 (.0076)*</td>
<td>0.0756</td>
<td>76441</td>
<td>0.192</td>
<td>£6,213</td>
</tr>
<tr>
<td>5. OLS Domain-weighted, excl. 2% (wave 15)</td>
<td>.1074 (.0154)**</td>
<td>-.0585 (.0231)**</td>
<td>0.2585</td>
<td>10124</td>
<td>0.545</td>
<td>£21,245</td>
</tr>
<tr>
<td>6. OLS Domain-weighted, excl. 2%, with personality controls (wave 15)</td>
<td>.0928 (.0147)**</td>
<td>-.0606 (.0218)**</td>
<td>0.3220</td>
<td>10114</td>
<td>0.653</td>
<td>£27,022</td>
</tr>
<tr>
<td>7. OLS Domain-weighted, excl. 2%, Income not logged</td>
<td>Y not logged</td>
<td>-.1472</td>
<td>0.0754</td>
<td>83129</td>
<td>£4,748</td>
<td></td>
</tr>
</tbody>
</table>

Note: ** p < 0.01, * p< 0.05, throughout. Unless stated income is logged, gross, current, household income. Mean income within the sample (using those in row 3) is £1273/month in 1987 prices or £29,323/year in 2005 prices. This is broadly in line with mean gross income based on the Expenditure and Food Survey 2005/6 (Jones, 2007), which estimates monthly gross mean income of £32,780 for 2005/6. The last row (7), showing the income compensation when income is not logged, is presented as a check.
Income compensations are based on the change in log of income necessary to balance out a change from no problem with street noise to a problem with street noise. This log income change is then converted to actual income at the mean income level.\footnote{Equivalently, this can be shown by returning to formula (2) in which we require the compensation income \( \text{compY} \) when the street noise dummy (SN) is 1 such that:}

\[
\text{SWB}_1 = \beta_1 \ln(Y) + \beta_2 SN_0 + \beta Z + \varepsilon = \beta_1 \ln(Y + \text{compY}) + \beta_2 SN_1 + \beta Z + \varepsilon
\]

\[\Rightarrow \beta_1 \ln(Y + \text{compY}) = \beta_1 \ln(Y) + \beta_2 (SN_0 - SN_1) \]

\[\Rightarrow \ln(Y + \text{compY}) = \frac{-\beta_2}{\beta_1} + \ln(Y)\]

\[\Rightarrow \bar{Y} + \text{compY} = e^{\frac{-\beta_2}{\beta_1} \ln(Y)} \quad \Rightarrow \quad \text{compY} = e^{\frac{-\beta_2}{\beta_1} \ln(Y)} - \bar{Y}\]

Table 40 shows a fairly wide range in valuations from street noise. Some judgement is therefore needed as to which is most appropriate, although compensation may be best understood as falling between a range of values (indeed values would be best presented with confidence intervals). Between 20 to 65\% of annual income is required to compensate for the impact on life satisfaction of street noise. This seems considerable. The fairly small income coefficient results in high value being attributed to most non-market goods that have a discernable impact upon SWB.

A further example can be seen in the case of valuing health. There is considerable interest in knowing how much a QALY is worth to society. This method can be applied to this context, using the Sf-6d available in waves 9 and 14 of the BHPS (see Chapter Three).

Life satisfaction is modelled as a function of net household income (current) and controls used in Chapter Three except all health variables are removed and replaced with the Sf-6d. As above, 2\% of income outliers are excluded. The implied valuation of a QALY is estimated from the income compensation required for a change in Sf-6d.

| Table 41: Regressions on life satisfaction and domain-weighted satisfaction, including sf-6d |
|----------------------------------|------------------|----------|-------|------------------|------------------|
|                                   | Income coeff.    | Sf-6d coeff. | R\textsuperscript{2} | Obs. | Equivalent units of log income for 0.01 change in QALY | Compensating income / year (2005 prices) |
| 1. Life satisfaction              | .1583 (.0186)**  | 3.4809 (.0631)** | 0.2308 | 20575 | 0.220 | £721,183 |
| RE                               |                  |           |          |      |                       |                             |
| 2. Life satisfaction              | .1653 (.0342)**  | 2.2819 (.1143)** | 0.0869 | 20575 | 0.138 | £434,095 |
| FE                               |                  |           |          |      |                       |                             |
| 3. Life satisfaction              | .1561            | 2.2483    | 0.0773   | 19737 | 0.144 | £454,270 |

\footnote{Equivalently, this can be shown by returning to formula (2) in which we require the compensation income \( \text{compY} \) when the street noise dummy (SN) is 1 such that:}
excluding long-term sick

<table>
<thead>
<tr>
<th>Description</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-value</th>
<th>p-value</th>
<th>Mean Income</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Domain-weighted FE, income not logged</td>
<td>.000397**</td>
<td>(.0001)**</td>
<td>2.2782</td>
<td>.1144)**</td>
<td>20577</td>
<td>£11,012</td>
</tr>
</tbody>
</table>

Note: This uses the same mean income as above for consistency.

Lines 1 and 2 of Table 41 compare coefficients on income and the Sf-6d using fixed and random effects. Controlling for fixed effects makes a considerable difference to the size of the Sf-6d coefficient. Given the strong relationship between personality and Sf-6d responses, controlling for individual heterogeneity seems essential in this case. The second row is therefore preferable; here a change in the Sf-6d from 0-1 is worth £434,095.

The controls used here include job status, one option of which is long-term sick which is likely to be highly correlated with health. Since we wish for all the health effects to be captured via the Sf-6d coefficient, controlling for being long-term sick may lead to the Sf-6d coefficient being underestimated. However, reducing the sample to only those in other employment states (row 3), shows minimal change in the health coefficient, but reduces the income coefficient (possibly due to excluding those for whom income is relatively important), therefore their inclusion gives the more appropriate valuation.

Using domain-weighted satisfaction as a dependent variable, in each case lowers the both the income and the Sf-6d coefficient, although overall fit of the model is improved.

The valuations vary greatly, but all cases where income is logged suggest a value far in excess of that which could usefully be applied in a policy context for health care allocation decisions. Addressing the potential endogeneity within the income coefficient may increase the size of this coefficient, hence reducing the compensations. However, even a doubling of the income coefficient would still result in compensations for the loss of a QALY that cannot usefully be applied to policy.

Cost utility analysis using QALYs has proved useful in supporting public resource allocation decisions within health care. QALYs focus only on health, but a more comprehensive measure of utility could be used as a generic outcome metric across a very broad range of public sector decisions. This would limit the need to value SWB changes in monetary units.
8.5 Conclusion

In order to allocate public resources to where they will do the most good, policy-makers need to know the changes in well-being brought about by changes in the type and level of non-market goods, such as health and the environment. Contingent valuation and SWB offer alternative methods for estimating the monetary value of non-market goods. However, the methods cannot be seen as different ways of assessing the same value but, rather, as two ways of perceiving what is of value. In the case of CV, the value of a good is given by our willingness to sacrifice other consumption for it \textit{ex ante}, whilst for SWB the value of that same good is given by how much it contributes towards our subjective assessment of how well our life is going \textit{ex post}.

Perhaps the biggest problem for incorporating SWB-income compensation values into the policy context is the fact that the coefficient on income has generally been found to be fairly small and even insignificant in some datasets (see Chapter Two). The low return on income, particularly for high-income groups, suggests that compensation in the form of a direct increase in household income is a very inefficient means to restore utility.

Whilst an income-based compensating variation may be high because, once beyond a subsistence level of income, additional income does not contribute much towards SWB, a resource-based compensating variation may lead to lower (and potentially more policy relevant) values for the non-market good, \( X \), we are trying to value. In this approach, the household would be given another non-market good, \( V \), up to the level at which it just compensates for \( X \). Of course, we would still need to know the cost of \( V \) but this could be estimated from the additional expenditure required, for example, the additional costs of providing health care.\footnote{There may be additional advantages in making compensation benefits in kind rather than cash if donor preferences are taken into consideration (Dolan and Olsen, 2002).}

If income compensations are large relative to average incomes due to the small income coefficient, this is not necessarily a problem for this method, as long as the income coefficient is un-bias and robust. However, very high valuations, whilst interesting, are problematic if they are not in-line with available resources or public expenditure priorities.

There are a number of issues that need to be addressed if robust non-market values are to be estimated from SWB studies. Currently, it is not possible to estimate robust income compensations, so we need to find robust ways of estimating the marginal utility of income or consider variants of cost-utility analysis. Since public money can be used on a range of non-market goods that, in various ways, seek to improve well-being, it makes good sense to value all goods in terms of their impact on SWB. This is in contrast to the private consumption trade-offs
that are expressed by WTP questions. Olsen and Smith (2001) argue that the appropriate cost within resource allocation decisions are the opportunity costs of different uses of public resources, since new programmes are likely to be funded out of existing budgets. Expressing all benefits and costs in terms of SWB would require large-scale studies that standardise the way in which SWB is measured and this is a major research challenge. Were benefits across different areas and departments of government to be expressed within a single SWB metric this would enable comparisons between alternative uses of resources to be made both within and between government departments.
Chapter 9: Conclusions, recommendations and future work

9.1 Recommendations for future work

Chapter Five concluded that SWB measures were sufficiently valid and reliable to contribute towards improving government policy. Despite this, they do not always meet the criteria for an ideal measure. Additional research aimed at a deeper understanding of these measures, and their failings, will help ensure that they are used and interpreted appropriately. Specifically, a greater understanding is needed on the processes involved when people respond to SWB questions, such that the meaning behind the numbers is understood.

9.1.1 Explore the extent to which SWB measures are interpersonally comparable

Chapter Five also concluded that interpersonal comparability of SWB measures was a reasonable assumption. However, there are some unresolved issues that challenge this assumption which warrant further exploration.

Some concerns relate to the use of the term ‘satisfaction’, and whether it is interpreted and treated differently across people. Chapter Three found that those with higher levels of education were less likely to respond at the top end of the life satisfaction scale. It was suggested that this may point to a difference in interpretation of the word ‘satisfaction’, potentially implying a limited imagination or lack of motivation. This could be explored by comparisons of evaluative responses across groups using different language and the same response scale (for example, ‘how satisfied are you with your life’, ‘how good is your life’, ‘how content are you with your life’, ‘how close does your life match your ideal’ etc.). This could shed light on whether a particular phrase draws out a different response according to group characteristics.

Other concerns relate to whether people use different implicit anchors when responding to these evaluative type questions. Anchors may vary from the ‘the best and worst life any human being is capable of experiencing and I am capable of imagining’ to ‘the best and worse life imaginable given my background, my current personal and socio-economic circumstances’. If an individual’s choice of implicit anchors is related to their characteristics, such as education, age or income, this will undermine interpersonal comparisons. Comparisons between different explicit anchors would also help understand the potential role of individual characteristics on how these questions are answered.
There is still much uncertainty surrounding the extent to which individuals recalibrate the SWB response scales in response to changes in their circumstances. For example, the presence of children may alter the use of the scale and the top of the scale may increase if life is better than previously thought possible. The change may not be restricted to the use of the measurement scale, but also to the concept itself, in that individuals may change their understanding of what it is to live a full and happy life.

If the measuring stick is not used in a consistent manner across time this undermines both inter and intra-personal comparisons. A considerable body of research has explored the issue of ‘reference shifts’ within assessment of health-related quality of life (HRQoL) (Albrecht and Devlieger, 1999; Allison et al., 1997; Breetvelt and van Dam, 1991; Carver and Scheier, 2000). Reference shifts are defined as “a change in the meaning of one’s self-evaluation of a target construct as a result of a) a change in the respondent’s internal standards of measurement (scale recalibration, in psychometric terms); b) a change in the respondent’s values (i.e. the importance of different domains which make up construct) c) a redefinition of the target construct (i.e. reconceptualisation)” (Sprangers and Schwartz, 1999: 1508). For example, individuals may change the comparison group that they use to set their anchor states following a change in circumstances. In the case of those newly diagnosed with a health condition reference groups may change from healthy individuals to other patients (Gibbons, 1999). Similarly, an individual’s expectations may change, in part due to the influence of their recent past experienced under the new circumstances.

Many of these health studies could be replicated looking at SWB measures. This would shed light on whether it is appropriate to directly compare SWB responses across changes in circumstances such as changes in health states, the presence of children or aging.

### 9.1.2 Explore the extent to which SWB measures are cardinal

Although there is reasonable evidence that the determinants of SWB are similar regardless of whether the life satisfaction or overall happiness scale is treated ordinally or cardinally, there is also some indication that the applicability of the cardinality assumption may be weaker at the top end of the scale (Layard et al., 2007).

Understanding the extent of cardinality is complicated by the fact that the determinants of SWB may have a different relationship at the top end of the scale than at other points in the scale (as seen in Chapter Three). This makes it very difficult to distinguish between genuine differences in the relationship between SWB and other characteristics and differences arising from inappropriately assuming cardinality.
Those responding at the very top end of the scale appear quite unusual, and would not necessarily be predicted from our understanding of what enhances SWB. For example, compared to those who respond at 6/7 those responding 7/7 have lower equivalised household incomes, are more likely to be widowed, more likely to be disabled and have problems walking, and more likely to have no education qualifications.96 This may point to a difference in scale use or anchor points within this group compared to others. Alternatively, complete satisfaction may require different types of circumstances. Greater understanding of the drivers behind extremely high responses would help clarify this issue.

9.1.3 Enable SWB to be used in economic evaluation

Economic evaluation requires a measure of benefit that can be compared against costs. Income compensations for SWB, as discussed in Chapter Eight, face theoretical and empirical concerns that are hard to envisage being overcome. Evaluation based on SWB directly is far more promising.

This requires a single SWB measure that uses language and a response scale for which it is most appropriate to assume interpersonal comparability and cardinality. For the evaluation of a factor which is expected to have a substantial impact upon SWB, the use of domain-weighted life satisfaction or a single evaluative question would seem likely to show sensitivity to the factor in question, as long as sample sizes were sufficient. Where the benefit is anticipated to be small for any individual, ESM or DRM type methods focusing on experiences in the moment and incorporating the additional consideration of time-use, may be necessary. These will necessarily exclude a longer-term evaluation of the individual’s life and have no means of incorporating a narrative quality to the assessment. If these factors are considered important then they would need to be accessed via alternative means, and the overall evaluation include non-commensurable benefits.

Similarly, the use of a single item evaluative question cannot tap into non-use values. A simple cost-benefit rule will be inappropriate where non-market goods touch on non-well-being issues such as beauty, knowledge, justice or morality. Where non-prudential values are seen as critical to the valuation other means of assessing public preferences and priorities towards these non-prudential goods will be necessary. Again, this will give a number of non-commensurable benefits.

96 For wave 14, equivalised net household income is £383 (€229) for those responding 7/7 compared with £456 (€269) for those responding 6/7; 33% have no education qualifications compared with 14%, respectively; 12% are widowed compared with 6%; 15% have problems walking compared with 7% and 11% are registered disabled compared with 5%.

270
Assuming only prudential values, the outcome could be summarised as net financial cost and net changes in SWB. A judgement will then be necessary on the value of changes in SWB. This judgement shares strong similarities with NICE’s efforts to derive a social value for an additional QALY.

At a theoretical level, a social welfare function that included everything of importance to society, including current and future well-being discounted at an appropriate rate for society, and non-well-being values, could be maximised subject to the constraint of available resources. This would then allow the marginal rate of return between alternative uses of resources to be equated. Despite the fact that trade-offs between different non-prudential values are implicitly made within decision making, an explicit social welfare function is unlikely to be derived such that it can aid in this kind of decision making, in part because many non-prudential values are not amenable to theoretical or practical measurement.

In the absence of a social welfare function, one option may be to explore the public’s WTP for a unit change of SWB. However, there is little reason to think that methods for extracting this figure would generate an accurate value. Alternatively, current decisions that are widely supported could be used as a touchstone for how much society values changes in SWB. For example, it may be possible to infer a value from currently supported medical or social care interventions that improve only SWB.

However, an explicit decision on the social value of changes in SWB may not be necessary. Policy makers can compare alternative uses of funds directly to each other, under a budget constraint of public willingness towards changes in rates of taxation, and the size of the economy.

Where an intervention involves both changes in current well-being and changes in length of life, it may be possible to express this in SWB terms under the assumption that everything of value in living is captured with the SWB measure. This requires considerable future work. It may be possible to rescale a SWB scale such that it uses a 0 (dead) to 1 (full SWB) scale, similar to a QALY. This would involve assessing the publics’ willingness to make trade-offs of improving SWB against additional years of life (TTO) or risk of death (SG), or including dead and perfect life on the SWB scale. This may be at the very bottom of the scale, fully dissatisfied, or at a mid-point in the scale, or indeed any point of the scale. For example, if the mid-point between being fully satisfied and fully dissatisfied is counted as a neutral point, in which the individual would consider they are indifferent between experiencing and non-experiencing, evaluations above this point add to aggregate SWB over a period of time, evaluations below this point subtract. This will rely on the assumption that the trade-off between length and quality of life is
constant at all levels of life expectancy. The incorporation of states worst than dead within QALY measurement has resulted in considerable additional complexity yet has been found to impact upon utility values (Franic and Pathak, 2003), and any assumptions about the location of dead within a SWB scale need validating.

As discussed in Chapter Two, aggregation across time is somewhat inconsistent with adopting a conception of well-being in which well-being is an individual’s evaluation of their life, since evaluation cannot easily be contained into one moment in time. Such inconsistencies leaves two options; not attempting to combine length of life and quality of life, or using a time-scale in which it is reasonable to assume evaluative aspects can be captured (e.g. one year) and accepting that the assumption that the evaluation relates only to this time period will be a theoretical oversimplification.

A problem may arise if the valuation of states changes with their duration as we require the state valued as equivalent to ‘non-existence’ to maintain the same valuation regardless of how long that state was experienced. Within HRQoL some severe states may receive a lower evaluation the longer is spent in that state implying a ‘maximum endurable time’ where patients tolerate a severe health state for a certain time beyond which their utility decreases making death a preferable option (Dolan and Stalmeier, 2003). Alternatively, some states may be valued more highly the longer is spent in then. For example, a 24 hour period may be evaluated more highly if it was only 23 hours excluding the 1 hour long bus ride in which boredom was the dominant experience. However, were boredom to be a permanent state it is less likely that this would be evaluated as worst than dead, since the individual would still attain other things important to their SWB such as contributing to their goals and ambitions and maintaining intimate relationships. Indeed, unless other attributes of value are incorporated into the momentary assessment it is unlikely that this can adequately be aggregated to present a combined quality and quantity of life assessment. The extent to which SWB scales measure only prudential value also undermines this exercise, as individuals have reasons for living beyond the SWB they experience.

Whilst establishing the trade-off between SWB and length of life is problematic and likely to rest on simplifying assumptions, research into public preferences towards trading off length of life and quality of life is likely to improve the trade-off used in evaluations compared to an implicit assumption, such as that used by Veenhoven’s ‘healthy life years’ (Veenhoven, 2000). This measure uses life expectancy multiplied by average SWB hence a percentage change in life expectancy is equally valued to a percentage change in SWB. Clearly there are benefits from incorporating life expectancy in judgements about the success of societies; however, we need at least some understanding of how changes in SWB are valued against changes in length of life.
If SWB researchers do not attempt to estimate the trade-off between length of life and SWB during life then interventions which impact upon both length and quality of life (health care and safety measures, for example) will require an implicit trade-off at the policy maker level, and reliance on existing measures. In the health care arena this would mean continuing reliance on QALYs, which do not capture everything of value nor every well-being change generated by health care policies. For example, although NICE guidelines recommend the inclusion of QALY changes in a patient’s family caused by treatment options, the impact upon families and carers may be better thought of as a well-being change rather than a HRQoL impact. Comparison of policies in terms of a measure incorporating all changes in well-being and all changes in length of life allow these effects to be considered, therefore should improve decision making.

9.1.4 Enable policy recommendations

There is still much that it not known about the determinants of SWB, particularly in terms of causality. Findings of association between a factor and SWB may lead to policy recommendations based on incomplete evidence.

Furthermore, any policy recommendations need to ensure that interventions are cost effective. Taking a simple example for illustrative purposes, if it is known that married people are happier (leaving aside for now the issue of whether this is an appropriate remit for government), this should not lead to policies to financially incentivise marriage unless other conditions are also met.

a. It is known that incentives will change behaviour and encourage some single people to marry and those who would have separated to stay together.

b. Those that marry or remain married due to incentives have higher SWB than they would have had otherwise, i.e. the causality is running from marriage to SWB regardless of the incentive for marriage.

c. The opportunity cost of offering the marriage incentives to all eligible people is less than the benefits that arise for the group of individuals who change their behaviour. This would also need to consider all individuals impacted by the marriage or the avoidance of its disintegration (including children), and the duration over which that loss occurs.

Even under circumstances where the benefits of an incentive policy appeared cost effective in terms of the SWB gained, compared to the alternative uses for those resources, society may also
wish to incorporate equity concerns into policy decisions. Support for married individuals who are happier than non-married individuals may further exacerbate the inequality of SWB. Were equality of SWB to be the main concern, additional assistance to single parents, who were shown in Chapter Three to score poorly across most measures of well-being, may be preferable to marriage incentives.

As this example shows, the evidence is currently too incomplete to lead directly to policy recommendations. Direct policy recommendations arising from the current SWB literature are limited. More information is needed about the effectiveness of policies (i.e. will any proposed intervention result in the anticipated change), and the cost of implementing policies.

The exploration of the relationship between social capital and SWB in Chapter Seven showed that, in this case, additional research is also needed to establish the direction of causality. This applies to many other potential areas where governments may wish to initiate policy. There is urgent need for more use of experimental and quasi-experimental studies to both clarify the direction of causality between the proposed determinants and SWB, and ensure that any proposed intervention is effective in initiating change.

Although the focus here has been on individual well-being rather than social well-being, policy recommendations require a society perspective. There is little understanding of public preferences towards maximising current well-being as oppose to other important objectives. Public preferences may give a certain weight to equality of well-being, beyond the maximising of well-being. The public may also care about particular changes in SWB more than others. Even if the life satisfaction scale is used cardinally at an individual level, public concern may rest more with changes below a level that they perceive as reasonable, so a change from 5 to 6 out of 10 may be given more social importance than a change from 9 to 10.

Non-prudential values (justice, aesthetic beauty, truth etc.) may also be weighed against maximising current prudential value. Similarly, the public may hold preferences towards increasing current well-being versus future well-being.

These type of preferences involve considerations beyond individual well-being into the field of social well-being, not focus of this thesis. However, greater understanding of social weights and efficiency-equity trade-offs is necessary before knowledge of individual SWB can be fully incorporated into the policy arena.
9.1.5 Improve understanding of the determinants of well-being

There are many areas where knowledge of the relationship between a particular factor and current and future well-being are limited. Future research should focus on understanding the most important factors in determining well-being, or areas which are critical for public policy, where intervention is likely to lead to the greatest improvement in well-being.

Psychologists have developed a range of theories to explain the determinants of well-being. For example, as discussed in Chapter Two, Ryan and Deci consider well-being to arise when the individual has their psychological needs met. They argue that competence, autonomy, and relatedness must be permanently satisfied for people to develop and function in optimal ways and achieve psychological growth, integrity and well-being (Deci and Ryan, 2000). There is a need to explore how these psychological factors relate to SWB. Whilst there are difficulties in investigating the impact of subjective variables, which may subject to endogeneity problems, understanding their influence on SWB is important.

Similarly, the evidence suggests that subjective interpretation of circumstances may be more important than the circumstances themselves, suggesting a need for greater understanding on what determines those perceptions. For example, perceptions of financial coping are strongly linked to SWB (as seen in Chapter Seven). It is therefore important to understand the circumstances that drive a poor perception of financial coping. For example, how does debt, lack of savings, lack of pension, material expectations, housing security or lack of financial self-control relate to perceptions of financial coping.

Since health is found to be highly influential upon SWB, it is important to know which type of health conditions most strongly impact upon SWB. These may differ from those ranked severe according to generic HRQoL instruments or preference-based QALY measures. The steady state SWB models should isolate those health states for which there is little long run adaptation, yet valuations of HRQoL, particularly those conducted from public preferences, are less likely to incorporate adaptation.

9.2 Other recommendations

9.2.1 Measuring well-being in the UK

At present SWB measures should be seen as work in progress, and resources dedicated to evaluating their performance and if necessary adopting future improvements. In their current form they still have a great deal to offer policy makers.
Governments and individuals will benefit from closely monitoring levels of SWB across society. At a minimum, monitoring should include overall life satisfaction, domain satisfactions and positive and negative mental health, at least annually.

The single life satisfaction question has been shown here to be sufficiently sensitive to identify policy relevant associations between SWB and personal, social and environmental characteristics. Given its efficiency at gaining SWB information, and the ease with which it can be included in self-report surveys, it should be widely used. As discussed above the exact wording, anchor points and number of response options requires further work before a clear recommendations on the most ideal format for this single question can be made.

In addition, domain satisfactions should also be collected. However, the exact domains and weights to be used is an area that requires further work in the UK context. The domains should represent a complete evaluation of an individual’s life. The example used in Chapter Three was based on data availability within the BHPS. Extensive work conducted on the PWI suggests important domains include the standard of living, health, what you are achieving in life, personal relationships, how safe you feel, feeling part of your community, your future security, and religion or spirituality (The International Well-being Group, 2006).

The PWI aims to be cross-culturally applicable, however, a problems arise when a domain appears to predict overall evaluation in one culture yet not another. For example, satisfaction with security is significantly associated with life satisfaction in some countries (and in some datasets) but not others (The International Well-being Group, 2005; 2006). As noted in Chapter Two the PWI uses domains such that they represent the first level deconstruction of the global question: ‘How satisfied are you with your life as a whole?’, and is simultaneously committed to the ability to judge domains both subjectively and objectively (The International Well-being Group, 2005).

However, these two aims may be in conflict. There may be important subjective aspects that do not have an objective counterpart. Satisfaction with our ethical behaviour, our capacity to deal with problems or our self-confidence, for example, are hard to envisage as objective indicators. There are other aspects of life that are predominantly subjective, such as satisfaction with levels of stress, satisfaction with amount of relaxation and peace, satisfaction with sleep quality and quantity, and satisfaction with energy and vitality. Whilst these may be validated with objective measures the benefit of doing so may be minimal if it is in their subjective form that they impact upon individuals’ lives.
Which domains should be included for the UK requires more consideration. Furthermore, issues relating to the substitutability between domains, whether particularly low or high values dominate the overall assessment, problems of reverse causality, and whether domains and weights differ across different cultural groups within the UK all require more exploration.

Nevertheless, replicating the PWI domains for the UK would be a promising start. Defra’s recent provisional addition of well-being indicators as part of the sustainable development indicators includes satisfaction with eleven aspects of life; relationships, accommodation, standard of living, local area, day-to-day activities, health, leisure, control, achievements of goals, future financial security and feeling part of the community. More work is needed to show that these domains represent a holistic picture of an individual’s life, and that there is no unnecessary replication of data collection.

Monitoring should also include tracking the numbers of people with a mental health problem, particularly numbers of those who are likely to benefit from treatment. As a society, we are likely to want to pay particular attention to those at the bottom of the SWB distribution, hence need to know how many people are currently suffering and the characteristics of those individuals so that policy can be targeted at addressing their suffering.

Whilst tracking those with particular mental health problems is important, such screening tools will provide limited information for the majority of the population. The mental health of the majority may be best monitored by instruments designed to focus on positive mental health, which incorporates aspects of subjective functioning. The Warwick Edinburgh Mental Well Being Scale (WEMWBS) offers potential for this and is currently under review and likely to be included in future Health Surveys.

In summary, monitoring well-being requires at least a single life satisfaction or overall happiness question, domain-weighted satisfactions which cover all aspects that are important to individuals, a positive mental health summary and a screening for particular mental health problems. The government should be aware of trends across well-being measures arising within particular subgroups, and inequalities in well-being measures across location, age, ethnicity, gender, sexuality, and socio-economic status.

Recommending a range of monitoring tools appears to undermine one of the criteria specified in Chapter Four of a single, preferably cardinal, measure of well-being. For economic evaluation and estimating the value of non-market goods, such a single measure is essential. However, the

---

97 Children should also be included in any monitoring of SWB. The same general structure could apply, but questionnaires require specific validation for each age group.
requirements for monitoring are slightly different. We wish not just to know the level of SWB of individuals, but also information about where problems lie, and in what areas of life individuals are unhappy. Although the single life satisfaction question is useful, it is unlikely to be sufficiently sensitive to illuminate everything of interest. Satisfaction across a range of domains helps to break down overall satisfaction to enable greater targeting of policy. For example, the 2007 Defra data on satisfaction with domains highlighted financial security (particularly for social class E) and satisfaction with community (particularly for social classes A and B) as areas where individuals were relatively dissatisfied (Defra, 2007).

The additional collection of mental health data will identify treatment needs. Monitoring positive mental health highlights particular areas of concern, and covers aspects of life that are unlikely to be picked up within domain satisfactions. For example, the Defra data that surveyed moods over the last two weeks identified restless sleep as a problem affecting 23% of respondents (Defra, 2007).

If a range of measures is collected, and they are not perfectly correlated, this raises the question of which measure is most accurate. At a theoretical level, positive and negative mental health and domain satisfaction can be perceived as inputs into an overall evaluation of life. However, since the overall evaluation may be picked up with some measurement error, it may not always be appropriate to defer to this value. Alternatively, we may think that all these measures are acting as proxies for some other conception of well-being.

At the monitoring level, using a number of different measures need not be of concern, since policy makers can be concerned with a range of different measures. Indeed, Defra’s sustainable development indicators currently contain many indicators all of which are considered important for the future of the UK.

Governments should be concerned with both current and future levels of SWB. Individuals are unlikely to fully incorporate an assessment of their future life within their assessment of SWB, even if the questions specifically request to do so. This may be in part due to lack of information and uncertainty about the future. Individuals are likely to draw heavily upon current levels of SWB, however, this may not be as good an indicator of future SWB as current levels of objective indicators.

Information on current and anticipated future well-being (either taken from the individual or estimated based on an understanding of the causes of SWB) require a trade-off to be made between current and future well-being. This may draw upon individuals’ time preferences. However, it may be considered that individuals adopt excessive discounting of their own future.
They may discount the future due to myopic preferences, partly through a lack of attachment to their future selves (Parfit, 1984) or the uncertainty that they will still exist in the future. This may be particularly so of children and young people.

The extent to which the government’s responsibility towards the future well-being of current citizens or future citizens including those who are yet to be born is greater than the preferences of current individuals towards their future self, and future (yet to be born) individuals, is a normative issue. However, governments will have at least some concern towards future well-being, therefore monitoring well-being necessarily includes monitoring current levels of well-being and the future causes of well-being. As such it will require monitoring of objective circumstances, and particularly those circumstances known to influence SWB.

### 9.2.2 Public policy

Many circumstances that determine an individual’s SWB can be influenced by both the external environment facing the individual and their personal choices. The provision of information on the determinants of SWB may help individuals make choices that are in their future best interests. For example, they may benefit from knowledge of research showing a relationship between SWB and involvement in the community, volunteering, long commutes, marriage and stable relationships, and hours worked. Frey and Stutzer (2007) note that this type of information should enable individuals to become better able to advance “what constitutes their idea of the good life, both individually and collectively” (Frey and Stutzer, 2007).

When the evidence permits, the government may wish to become directly involved in establishing conditions in which individuals can lead happy lives. This requires a framework for deciding upon the appropriate remit for government involvement. Although there may be universal agreement that higher SWB is better than lower SWB this does not necessarily generate sufficient justification for government involvement.

Economic theory offers a framework for justifying government involvement in areas of market failure and incomplete markets. This may include areas where individuals have no opportunity to express their preferences (e.g. incomplete capital markets) or where their current preferences are perceived to be ill informed such as in the case of merit goods (e.g. education, health). Markets may also fail due to externalities (e.g. noise pollution, crime), public goods (e.g. parks, policing) or natural monopoly (e.g. public transport). Furthermore, government intervention may be justified on the basis of the provision of basic necessities, such as housing and water.

Many of the factors that have been found to relate to SWB fall into areas where there is accepted market failure, and in many cases, existing national and local government
involvement. This includes health, education, housing policy, environment policy, supporting communities, noise pollution, crime and access to public transport. There is therefore a clear role for using SWB measures, among other outcome measures, to ensure that current government intervention is as efficient as possible. As noted above, the evidence, particularly in terms of cost effectiveness of interventions, does not yet allow clear policy recommendations to be drawn. However, the evidence is sufficiently strong to justify future research into these areas with a view to establishing whether local or national government policy could effectively enhance SWB via these determinants.

A strong relationship to SWB has also been found in areas that are not traditional remits of government, nor are perceived to suffer from market failure. If an attribute appears to be under-consumed, in the sense that the evidence suggests that were the individual to consume the attribute their SWB would be enhanced, does this make it a merit good? There is some evidence that happy people contribute more to society, but this is unlikely to be sufficient to justify that the social benefit would exceed the private benefit. The case of merit goods therefore requires some malfunction of individual preferences. In the case of education it is accepted that young people are unlikely to have preferences that adequately incorporate their future self or future states of the world, and hence are not able to make informed decisions. The individual’s assessment of benefit is likely to be underestimated, making the externality apply to the individual, either now or in the future.

Government intervention seeks to change or override some preferences that are not in the individual’s best interests. Where the problem can be framed as an externality imposed upon the future individual, it is reasonable that the government treat that individual as a future member of society, regardless of the weight the individual attributes towards their future self. Consequently, intervention for pensions, safety, or healthy lifestyles can be justified upon those grounds.

There are areas within the SWB literature in which government policy has a less obvious role, specifically marriage and religion. Intervention would require that preferences were poorly informed, and it is hard to imagine confidence in such a claim, particularly in the case of religion where the truth of a belief system cannot be proven.

Although specific policies cannot yet be drawn from the SWB evidence, it is possible to conclude that income (at high levels of income) is only a small element within an individual’s utility function, and part of that income effect is likely to arise through changes in income rank. This knowledge requires a shift in macroeconomic policy away from maximising economic growth. Growth may still be instrumentally beneficial for well-being in terms of provision of
employment, support for the public sector, maintaining a stable, low inflation and low unemployment economy and supporting social and political stability. Growth cannot, however, be treated as valuable in its own right. For example, economic growth derived from capital intensive extraction of natural resources where the income remains in the hands of those with low marginal utility of income is unlikely on aggregate to show any well-being improvement, and may result in a decline in aggregate well-being.

A shift in thinking from the assumption that economic growth is beneficial *per se*, to perceiving economic growth as beneficial only for the attributes it delivers for society should enable a greater focus on how growth translates into factors that improve well-being. For example, employment is important, but so is the quality of that employment.

Income taxation policy is another example where the SWB evidence should be brought to policy decisions. Although Chapter Six concluded that the evidence on relative income effects is as yet too uncertain to be appropriately incorporated into taxation policy. The combination of some relative income effect, and moderate or small impact of absolute income, does suggest that any increase taxation, particularly on increases in income, will have a minimal impact upon SWB.

There is nothing new in the recommendation that economic growth should not be treated as a sufficient outcome measure. For example, Sen has argued that, "A person's well-being is not really a matter of how rich he or she is . . . commodity command is a means to the end of well-being, but can scarcely be called the end in itself" (Sen, 1985a: 28). However, the SWB data offers an increasing body of evidence to lend the recommendation both weight and urgency.

### 9.3 Extensions and improvements to the data analysis

In Chapter Three the use of the generalised ordered probit model helps to illuminate the unusual relationship between education and life satisfaction. The results from this model suggest that this type of analysis may also raise new understanding across other well-being measures.

Chapter Three focuses upon preference satisfaction and mental state measures only, and does not incorporate any substantive good measures. There are some measures available which adopt a needs based approach to well-being (e.g. Ryan and Deci’s psychological needs scale) and there is also on-going work to develop a complete measure of well-being from both a needs based (e.g. Cox and Ajetunmobi, 2007) and a capabilities perspective (e.g. Coast *et al.*, 2006; Anand and van Heeds, 2006). When data becomes available on these new measures it will be beneficial to incorporate them into the type of analysis conducted in Chapter Three. Many of the criteria for an ideal measure of well-being can not be adequately addressed for substantive good
measures (see Appendix III) because of a lack of empirical data. Once data is available on substantive good measures it will be possible to test their validity.

It would also be beneficial to see whether the findings from the comparisons within Chapter Three are replicated within other datasets. For example, The English Longitudinal Study of Ageing (ELSA) contains both the CASP-19 and the GHQ-12.

The control variables used in Chapter Three were drawn from those commonly used within the literature. However, a recent study on the GSOEP (Luechinger, 2007) also controlled for (and found significant) whether this is the first or second year the interviewee has been in the panel. This is serving as a proxy for interviewing experience and panel learning effects (d'Ambrosio and Frick, 2004), thus would have been a useful control because the analysis in Chapter Three is not on a balanced panel, and compares different scales that have been incorporated within the BHPS for varying lengths of time. Consequently, interviewee experience with the scales may not be generating random error.

The relationship between SWB and income is relevant to all the data analysis, but particularly Chapters Six and Eight. Concerns are raised in the thesis about the accuracy of the income measure, possible reverse causality, possible omitted variables which may be related to income, and the considerable change in magnitude of the income coefficient when this endogeneity is addressed through the use of instrumental variables. The data analysis would therefore benefit from reconsideration in the light of instrumenting income. This may either use data within the BHPS, or data that can be tied to households in the BHPS either through individual characteristics or geographical area.

Variables which have been used by other studies to instrument for income include work experience, regional unemployment rate, parental education, spouses parental education, and work experience when looking at health (Ettner, 1996); interest rates and stock prices when looking at the permanent income hypothesis (Campbell and Mankiw, 1990); data on quits, promotions, layoffs, hours ill and unemployed when looking at the relationship between income and consumption (Altonji and Slow, 1987); and 10-year average income when looking at charitable giving (Randolph, 1995). Others have used consumer durables such as access to car(s), type of accommodation or ownership, lagged income or lagged values of other productive variables. The particular difficulty with the SWB-income relationship is that there are multiple routes in which many of these variables can impact upon SWB.

Although Chapter Six finds a negative relationship between past income and SWB when controlling for current income, additional analysis did not support that this is derived from a
straightforward annual adaptation. One explanation raised is that adaptation occurs over a longer period of time and may relate to income during childhood. This could be explored within the BHPS through the youth panel. As the sample size of those respondents with detailed income data from their youth expands this dataset will be ideal for exploring such issues.

Chapter Seven finds that the perception of income is a strong determinant of SWB. This warrants much greater exploration. It is important to know whether income satisfaction is drive by income expectations, or perceptions of income and financial coping, or additional needs and low assets. More generally all models would benefit from greater control of individual and household assets and wealth.

Chapter Six finds that although reference income is important, the regional average income is not a very robust determinant of life satisfaction. The literature suggests that the impact of reference income may vary depending on closeness of the reference group. This could be explored with other datasets or potentially through the BHPS by considering smaller regional areas which could be linked to census data.

The impact of household and individual income of siblings was considered in earlier work. This is not reported here as no significant relationships were found. This may be in part because those for whom sibling income could be traced within the BHPS were those who lived with a sibling earlier in the dataset, hence were fairly young and less impacted by relative income. Consequently, as the dataset expands it may be productive to return to this analysis.

Chapter Seven identified contact with neighbours as important to SWB. However, the problem was also raised that the quality of contact could vary considerably within these contacts. Since intimate, family and community relationships have been shown to be strongly associated with well-being it is important to find a means of exploring the impact of the quality of these relationships.

Chapter Seven found that when instrumenting for contact with neighbours the size of the coefficient increased considerably, which may suggest measurement error within the talking to neighbours variable. It would be interesting, therefore, to find more objective means of measuring this variable, such as actual numbers of neighbours that the individual can name. This would also enable other possible causes of endogeneity to be explored, for example, is there any relationship between accuracy of reports of the frequency of talking to neighbours and reports of SWB.

One of the unexpected findings of Chapter Seven is the negative relationship between length of time in the neighbourhood and talking to neighbours. Provisional exploration found that this
variable is correlated with many other variables in the data, but was unable to find an adequate explanation for this somewhat counter-intuitive finding. Greater understanding, possibly arising from qualitative work, would be useful to this analysis.

Another factor that was not considered within the analysis in Chapter Seven is the impact of homogeneity of neighbourhood, in terms of characteristics such as race, age, income and socio-economic status. This is an area where any short-term well-being implications would need to be carefully weighed against both long-term implications and non-prudential concerns.

9.4 Conclusion

Throughout this thesis it has been emphasised that WTP cannot be used as metric of value or well-being. The price an individual is willing to pay does convene some useful information, namely how much the individual is willing to sacrifice in terms of money, at a particular moment in time. This may still be some indication of strength of preference, it is simply not a measure of ex-post change in well-being, either at an individual level and certainly not on aggregate once relative income effects are taken into consideration. The use of SWB as a metric of value offers considerable advantages over WTP.

As a discipline focused on the efficient use of resources, economics has systematically failed to provide the tools to help policy makers use their resources efficiently, such that well-being (and other social concerns) can be maximised. This thesis has shown that this can, and should, change.
References


References


References


References


References


References


References


References


References


Lelkes, O. (2002) Tasting freedom: happiness, religion and economic transition. CASEpaper 59, LSE.


References


Martin, M. & Westerhof, G.J. (2003) Do you have to have them or should you believe you have them? Resources, their appraisal, and well-being in adulthood. *Journal of Adult Development*, 10, 99-112.


Powdthavee, N. (forthcoming) Putting a price tag on friends, relatives, and neighbours: using surveys of life satisfaction to value social relationships. Journal of Socio-Economics


References


References


Wildman, J. and Jones, A. (2002) Is it absolute income or relative deprivation that leads to poor psychological well being? A test based on individual-level longitudinal data. YSHE, University of York.


Appendix I: Background on the BHPS

The British Household Panel Survey (BHPS) is a longitudinal survey designed to capture information on a nationally representative sample of the population of Great Britain that began in 1991. It was extended to cover Northern Ireland in 1997. It has been developed and run by The University of Essex, Institute for Social and Economic Research. The BHPS is a sample survey and as such subject to both sampling error and bias due to non-sampling error (principally non-response and attrition of the original panel).

The BHPS surveys each adult (16+) member of a nationally representative sample of about 5,500 households, making a total of approximately 10,000 individual interviews rising to about 15,500 by wave 15. The same individuals are re-interviewed in successive waves and, if they leave their original households, they are re-interviewed along with all adult members of their new households. From wave four onwards a youth survey of 11-15 year old household members was included. Babies born to Original Sample Members (OSMs) and new members of original households are also included in subsequent waves. Thus the sample aims to remain broadly representative of the, non-immigrant, population of Britain (Taylor et al., 2004; Nathan, 1999).

In wave one, eligibility depended on domestic residence in England, Wales, or Scotland south of the Caledonian Canal. In subsequent waves OSMs were followed into institutions (unless in prison or in circumstances where the respondent was not available for interview e.g. too frail, mentally impaired etc.) or into Scotland north of the Caledonian Canal.

From wave seven, the BHPS starting providing data for the United Kingdom European Community Household Panel (UKECHP). The BHPS consequently incorporated a sub-sample of the original UKECHP, including households still responding in Northern Ireland, and a ‘low-income’ sample of the Great Britain panel. In wave nine additional samples were taken in Scotland and Wales to boost the relatively small Scottish and Welsh sample sizes, so that country level analysis could be undertaken.

Interviewing begins in September and continues into the Spring, with many questions relating to the year to September. The mode of data collection moved to Computer Assisted Personal Interview (CAPI) in wave nine, aiming to minimise missing data and reduce the level of data cleaning and editing.

98 For information about the BHPS see http://www.iser.essex.ac.uk/ulsc/bhps. For details of using the BHPS see Lambert (2006).
The BHPS administrators carry out a number of procedures to maintain contact with respondents between waves, such as spending out thank you letters, a gift voucher and change of address card. The longitudinal response rate of the OSMs for the second wave was 86% falling to 55% in wave 13. Some of those dropping from the main sample are given a shortened interview by proxy and some via telephone.
## Appendix II: Surveys referred to in the text

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL</td>
<td>Americans’ Changing Lives</td>
</tr>
<tr>
<td>BHPS</td>
<td>British Household Panel Survey</td>
</tr>
<tr>
<td>ESS</td>
<td>European Social Survey</td>
</tr>
<tr>
<td>GSOEP</td>
<td>German Socio-Economic Panel Survey</td>
</tr>
<tr>
<td>NCDS</td>
<td>National Child Development Survey, US</td>
</tr>
<tr>
<td>NSFH</td>
<td>National Survey of Families and Households, US</td>
</tr>
<tr>
<td>RLMS</td>
<td>Russian Longitudinal Monitoring Survey</td>
</tr>
<tr>
<td>SCB</td>
<td>Social Capital Benchmark</td>
</tr>
<tr>
<td>US GSS</td>
<td>US General Social Survey</td>
</tr>
<tr>
<td>WVS</td>
<td>World Values Survey</td>
</tr>
</tbody>
</table>
Appendix III: Substantive good accounts of well-being

III.1 Objective list or substantive good accounts

Objective list or substantive good concepts of well-being provide a set of attributes which are thought to constitute well-being. These may include such attributes as “moral goodness, rational activity, the development of one’s abilities, having children and being a good parent, knowledge and the awareness of true beauty” (Parfit, 1984: 499).

Below distinguishes between four types of substantive good accounts, basic needs, perfectionist or flourishing, capabilities and list accounts. These all adopt a similar format of listing a set of attributes which are important for exploring issues of well-being, however, they have important structural differences. Any of these accounts may be categorised within the last category of list accounts, however, the first three hold specific characteristics.

Needs accounts are distinct because, although they give a set of attributes, they do not identify well-being with that list but rather see the list as essential inputs to attaining well-being. Perfectionist accounts are distinct because they derive the attributes on the list from a particular perspective, that of attainments necessary to live a perfect human life. Capability accounts are distinct because of their focus on whether the individual has the opportunity to attain attributes of value (as with needs accounts the source of that value does not need to be specified).

Other list accounts do not necessarily provide any rationale or formal theory of why an attribute constitutes well-being, instead calling upon intuition, religion, social tradition or public consensus.

III.1.1 Needs accounts

III.1.1.1 Theories of need accounts

Needs accounts assert that the only politically relevant aspect of well-being is whether essential human needs for survival and flourishing are met. However, needs accounts are not formal accounts of well-being but rather suggestive of attributes which are the essential means for creating well-being. Needs theorists do not agree on a single list of human needs. To represent the broad spectrum of this material I shall briefly outline prominent versions by Rawls (1973), Doyal and Gough (1991), Max-Neef (1992) and Ryan and Deci (2001).
Appendix III: Objective list or substantive good accounts

Some needs accounts focus on basic needs necessary for attaining well-being. For example, Rawls (1973) developed an index of primary goods which include ‘rights, liberties and opportunities, income and wealth and the social bases of self-respect’. However, Rawls still held that well-being is determined by the satisfaction of rational desire (Rawls, 1973: 92-93). This conception therefore implies that to have high well-being an individual must have adequate primary goods which are the means for satisfying the individual’s rational desires, whatever they may be.

Doyal and Gough drew up a list of basic human needs or universal goals, which they intended to be objective and non-culturally sensitive, using the criteria of that which is necessary to avoid serious harm and be an effective member of one’s society (Doyal and Gough, 1991). They specify two basic needs; physical health and autonomy (which is seen as the capacity to initiate an action through formulating aims and beliefs) and 11 intermediate needs. These are seen as pre-conditions for well-being.

Max-Neef also constructed a matrix of universal human needs (Max-Neef, 1992). Well-being here depends on the possibilities people have to adequately satisfy their fundamental needs. The nine elements of the matrix are subsistence, protection, affection, understanding, participation, leisure, creation, identity and freedom.

Whilst earlier basic needs accounts focused on physical resources necessary to lead a good life, recent versions have emphasised the psychological resources necessary for well-being. Ryan and Deci (2001) posit three basic psychological needs - autonomy, competence and relatedness - and argue that fulfilment of these is necessary for well-being.

III.1.1.2 Measuring needs accounts

Needs theorists set out those attributes in a person’s life that are essential for attaining well-being. Although meeting needs focuses upon inputs to well-being, satisfaction of needs may serve as an equally good proxy for well-being as a focus on outcomes, and have the advantage that the formal theory of well-being is left open.

Needs accounts do not necessarily deliver a method to establish the level of individual well-being, rather they point to what needs to be in place for well-being to arise. For example, Max-Neef’s matrix of human needs is used to enable communities to identify ways in which their resources, cultural traditions, institutions etc. inhibit them from meeting their fundamental needs, rather than to give a measure of well-being. The Max-Neef matrix has recently been

---

99 Where harm is defined as a disablement in the pursuit of one’s vision of the good.
translated into a 51-item well-being questionnaire for use in Dumfries and Galloway, although this is still in the development phase (Cox and Ajetunmobi, 2007).

Instruments which have been developed from a needs perspective are Ryan and Deci’s Basic Psychological Needs Scales (Deci and Ryan, 2000). These are a family of scales measuring individual psychological needs of competence, autonomy and relatedness. Basic needs accounts that incorporate psychological characteristics, such as autonomy, lend themselves to subjective rather than objective judgements. It is not just the case that self-report has practical advantages, but rather that certain feelings are seen as an important component of basic psychological need.

III.1.2 Perfectionist or flourishing accounts

III.1.2.1 Theories of perfectionist or flourishing accounts

Perfectionist theories conceive every living being as having an essence which sets the goal for the life of that being. How well an example of that species’ life goes depends upon how closely the ideal of the species is met, or whether it ‘perfects’ the nature of that species. Perfectionism therefore identifies the good by reference to human nature. If knowledge, achievement and friendship are good it is because they realize aspects of human nature, independently of any pleasure they may bring (Hurka, 1993). Perfectionist theorist identify with the concept of well-being as the ‘good life’, rather than the life which is good for the individual.

Most perfectionist accounts of well-being draw from Aristotle’s understanding of eudaemonia (which has been variously translated as happiness, well-being, living well, good fortune, flourishing, and fulfilment) as the state that all fully rational people would strive towards. Aristotle argues that because man is both social and rational, eudaemonia must consist of attaining both social and intellectual goals (Nicomachean Ethics, 1,7,13). He offers two candidates for eudaemonia - contemplation (intellectual activity aimed at attaining the truth) and virtuous activities (such as the acts of politicians or soldiers, or activities involving love and friendship). Pleasure is distinct from eudaemonia, but it is seen as usually arising from activities which constitute eudaemonia.

However, versions of perfectionism vary. For example, Hurka presents human nature (and therefore human good) as physical perfection and theoretical and practical reason (Hurka, 1993).

Psychologists have tended to move away from species perfectionism into individual attainment, and an individual reaching their true potential. For example, Ryff and Singer present well-being
as “the striving for perfection that represents the realisation of one’s true potential” (Ryff, 1995: 100).

### III.1.2.2 Measuring perfectionist or flourishing accounts

Measuring well-being according to flourishing accounts is difficult because all chosen attributes are important in their own right, and therefore cannot be easily traded-off against each other. Two scales which attempt to give an aggregate score for well-being from an individual flourishing perspective are Ryff’s Psychological Well-being Scale (PWBS, Ryff, 1989) and Peterson’s Orientations to Happiness Scale (Peterson et al., 2005). Ryff’s Psychological Well-Being Scale measures aspects of human actualization: self-acceptance; positive relations with others; autonomy; environmental mastery; purpose in life and personal growth. The Orientations to Happiness Scale is based on a belief that the ‘good life’ is made up of three aspects: the pleasurable life, the meaningful life, and the engaging life (Seligman, 2002). The engaging life dimension incorporates the benefits arising from absorption in tasks. Csikszentmihalyi (1990) describes this state as "flow", and presents it as beneficial even when not linked to pleasure.

The European Social Survey (ESS) 2006 module on well-being incorporates functionings from a flourishing perspective (such as autonomy, competence and interest in learning) but also includes feelings and evaluations (such as optimism, life satisfaction and positive affect) (Huppert et al., 2006). Functionings are taken as inherently important, independent of any subsequent impact upon personal feelings.

### III.1.3 Capabilities accounts

#### III.1.3.1 Theories of capabilities

Amartya Sen has developed an interdisciplinary framework, known as the capability approach, which can be used for analysing issues of poverty, inequality and well-being (Sen, 1980, 1984). The capability approach uses the notion of basic capabilities, which are abilities or opportunities to achieve certain functionings, such as being a certain type of person or doing certain things (e.g. moving about, meeting one’s nutritional requirements (Sen, 1980)). An individual’s capability represents the alternative combinations of functionings people can achieve, from which they can choose one combination (Sen, 1993). Capability is the ability or capacity to achieve different states or conduct different activities. Hence the theory focuses mainly on access rather than the actual outcomes. The actual functioning attained is taken as an individuals ‘well-being achievement’, whereas the options available is their ‘well-being freedom’.
“Capability is thus a kind of freedom: the substantive freedom to achieve alternative functioning combinations (or, less formally put, the freedom to achieve various lifestyles)” (Sen, 1999: 74-75).

### III.1.3.2 Measuring capability

Sen has stressed the belief that various capabilities are incommensurable and cannot be measured and compared by some common, deeper, measure such as utility.

Sen has avoided setting out exactly what constitutes human functionings, beyond giving some examples, such as longevity, health and self-respect, preferring to leave the capabilities account as a theoretical structure within which welfare issues can be analysed. Sen (1993) argues that this provides more choice for individual societies to determine for themselves important functionings, thereby ensuring the relevance of the approach across different cultures. Sen notes that the incomplete capabilities theory is, “consistent and combinable with several different substantive theories” (Sen, 1993, 49) which may be filled in by reasoned agreement. However, priority is given to the capabilities and functionings that we have ‘reason to value’, and which arise from democratic public deliberation (Sen, 2004). The outcomes upon any list of capabilities and functionings are those that would arise from informed and well-reasoned desires, suggesting an idealised desire satisfaction account underlying the capabilities approach. However, how this could be operationalised remains unclear. As observed by Sugden:

> “Given the rich array of functionings that Sen takes to be relevant, given the extent of disagreement among reasonable people about the nature of the good life, and given the unresolved problem of how to value sets, it is natural to ask how far Sen’s framework is operational” (Sugden, 1993: 1953).

Martha Nussbaum does develop the capabilities account into a more complete objective-list which draws upon the capabilities necessary to achieve functionings which comply with Aristotelian notions of well-being as living a truly human life (Nussbaum, 2000). She argues that essential flourishings necessary for people to be considered as living well and in dignity are universal across cultures. These basic capabilities are seen as having value in themselves regardless of whether they are desired and perceived to be valuable by the individual. They are derived from those attributes which are seen as important across society and attained by “overlapping consensus” among people around the globe and well-being literature.

Her list of features essential to a full life consists of ten dimensions including life, bodily integrity, practical reason, affiliation and play (Nussbaum, 2000). This list is aimed at helping political institutions develop constitutional guarantees that ensure every citizen has access to the

---

100 ‘Overlapping consensus’ is a term coined by Rawls, meaning compatibility with a range of metaphysical views of the world, different ethical or religious views and different views of human nature.
necessary circumstances for good functioning. This does not therefore imply that a similar list would be derived if the purpose of the list was to measure well-being, or that this list gives a complete measure of well-being. Nevertheless, Nussbaum’s list is being used as a basis for deriving capability set which is designed to measure well-being, as discussed below.

The British Equalities Review recommends the use of a capability approach within public bodies to evaluate progress towards equality. They have drawn up a list of capabilities derived initially from international human rights and refined through democratic deliberation and debate. The list of ten domains includes the capability to be knowledgeable, to participate in decision making and to enjoy individual, family and social life (Vizard and Burchardt, 2007). This approach extends the work of Nussbaum from an exercise in understanding what should be constitutional guarantees to monitoring inequality which requires measurement. Inequality is presented as irreducibly multi-dimensional hence not suited to a single index. Routine monitoring is recommended by focusing on one or more particularly salient aspects or ‘spotlights’ within each domain. These are to be supplemented by ‘roving spotlights’ which are varied year to year, highlighting important aspects rather than representing a summary of inequality within the domain as a whole (Burchardt and Vizard, 2007: 21).

Suggested spotlight or headline indicators include life expectancy, victimisation of violent crime, self-reported health status, income, social isolation index and self-esteem. Hence, although the approach is one of capabilities, measurement focuses predominantly on outcomes. Indeed, the practical difficulty of measuring true opportunity is likely to make this the case in most attempts to measure capability. Burchardt and Vizard (2007) note that disadvantage across a whole subgroup or across a number of domains for a subgroup is likely to point to a failure in capability rather than expression of choice. Additional assessment of the perceived degree of choice and control to bring about outcomes is also recommended to supplement information on outcomes.

This list of capabilities was not designed to assign a level of well-being to each individual, rather to shed light on existing inequalities and monitor progress towards their elimination. Nevertheless, there must be cross over because if any factor that mattered to individual well-being was not included on the list for inequality, we would have to be concerned about inequality in that domain.

Anand and van Hees (2006) have developed a questionnaire designed to pick up both capabilities and achievements. This focuses on seven domains: happiness, achievement, health, intellectual stimulation, social relations, pleasant environment and personal integrity. The questions aim to tap into ‘capability’ by asking about perceptions of ‘scope for achievements’
Appendix III: Objective list or substantive good accounts

(e.g. I feel the scope to achieve things in my life is: very good (1) to very inadequate (7)), and ‘achievements’ by asking about satisfaction with, or presence of, achievements (e.g. I have satisfying social relations: strongly agree (1) to strongly disagree (7)). The authors note that this measure will identify only whether an individual thinks that they have opportunities to achieve things whether or not they actually could.

Anand et al. (undated) also developed a 65 question survey designed to capture all the items on Nussbaum’s ten dimensional list (questions include: I appreciate and value plants, animals and the world of nature; My idea of a good life is based on my own judgement). Although an aggregate figure is not given, these capability responses are regressed against life satisfaction, implying by functional form that trade-offs between capabilities are possible.

There is also ongoing work to develop a capabilities measure for older people that adopts a multi-attribute questionnaire with a similar structure to the EQ-5D which may have the potential to value capability level from a scale of dead to full capability (Coast et al., 2006; Grewal et al., 2006).

As yet there has been little evidence to show that measures designed to pick up capabilities offer a level of individual well-being which is distinct from a measure designed to pick up functionings. However, this research is still in its infancy.

III.1.4 List accounts

III.1.4.1 Theories of list accounts

List accounts offer a list of attributes that are taken as constituting well-being. They do not necessarily provide a formal theory of well-being although some may present attributes as having inherent properties which make them good for the individual. The content of lists vary but they tend to include items such as economic resources, political freedom, good health, knowledge and literacy.

For example, John Finnis claims that life, play, aesthetic experience, sociability, practical reasonableness, knowledge and religion are intrinsically good for people irrespective of whether they are pleasurable or whether individuals have desires for them (Finnis, 1980).

“It is obvious that a man who is well-informed, etc., simply is better off (other things being equal) than a man who is muddled, deluded, and ignorant, that the state of the one is better than the state of the other, not simply in this particular case or that, but in all cases, as such, universally and whether I like it or not. Knowledge is better than ignorance.” (Finnis, 1980: 72, emphasis in original)
Some lists present justification for why elements are present. For example, in Grisez, Boyle and Finnis’s (1987) list of ‘basic human goods’ inclusion of an item on the list is justified if it is desired for its own sake, in other words it provides reasons for decisions and action, and can be reasoned through repeated questioning of ‘why I / others do what we do?’ Grisez et al. (1987) suggest a list of basic reasons for actions that includes life itself, knowledge and aesthetic experiences.

In principle there is no limit to what someone may claim should be included in a list of well-being. It may include both objective states and subjective elements, such as mental states. Veenhoven (2000) offers an example of a list which combines living conditions, living skills (the capacity to deal with the environment) and outcomes of life (having a long, healthy and happy life).

### III.1.4.2 Measuring list accounts

From a list account of well-being an individual’s life goes better if they attain more of the attributes which are held to constitute well-being. More of any attribute upon the list, when others are constant, is always assumed to represent an increase in well-being. If these attributes are treated as fully non-commensurable, except under circumstances where one attribute increases while others remain constant, levels of well-being cannot be compared. If trade-offs are allowed, it may be possible to derive a level of well-being according to the balance of attainments. These trade-offs may not be at constant rates, but change as levels of a particular attribute decrease – to the extreme of one attribute acting as a constraint upon the level of well-being.

Many list accounts of well-being are operationalised via social well-being rather than individual well-being (see Sharpe, 1999 for a review), but adhere to the idea that different items are non-commensurable. Some examples include the Audit Commission in England and Wales quality of life indicators for local authorities, and the UK Government Sustainable Development indicators.

Well-being from a list perspective at the individual level can be measured by surveys such as the Swedish Level of Living Surveys (Duncan et al., 2005: 12). ‘Level of living’ was defined as “an individual’s command over resources... through which she can control and consciously direct her living conditions (Johansson 1970: 25, cited in Duncan et al., 2005: 12 (original in Swedish)). This implies a similarity with the capabilities framework, since it considers the scope an individual has for action. The survey adopted a multi-dimensional approach with nine domains including family and social integration, economic resources, health and access to
Appendix III: Objective list or substantive good accounts

health care and schooling and education opportunities. Whilst this survey was originally designed to measure objective aspects of quality of living, such as the presence of hot water within a house, it has recently incorporated subjective indicators including questions on overall life satisfaction, subjective health state, satisfaction with changes, and income satisfaction (Duncan et al., 2005: 16).

There may be some theoretical ambiguity within lists accounts of well-being as to whether the attributes constitute the outcome of well-being because they have properties of goodness, or whether the attributes are sources of well-being. Needs based accounts clearly focus on the latter. However, substantive good or objective lists for clarity of theory should be outcomes only, or those attributes that non-instrumentally improve an individual’s life. However, in practice many measures drawn from this perspective incorporate both inputs into well-being and outcomes.

As seen in the Swedish Level of Living survey and Veenhoven’s list they can incorporate subjective elements. These may include overall happiness or life satisfaction as additional items on the list or consider individual perception of quality of an attribute (such as housing) as an important indicator of the achievement of a domain. Indeed, lists may incorporate both an individual’s achievement within a domain measured subjectively and measured objectively, with the potential to give a combination of a subjective and objective assessment of the each domain.

The time frame over which lists and measurement instruments discussed here identify well-being is often somewhat unclear, all appearing to refer to some duration of time rather than the current moment or complete life. Lifetime well-being is captured when length of life is incorporated within the list, where it is treated as an additional, non commensurable good.

Different substantive good accounts start from different perspectives and justification for including an item as part of well-being may differ, they also vary in purpose, some focusing on minimum requirements for a decent life, others on human flourishing. That said, the content of lists are often very close (Alkire, 2002). However, there are some exceptions, for example Nussbaum’s inclusion of ‘play’ and ‘the ability to live in a fruitful relationship with animals and the world of nature’ are unique to her list of capabilities. That these should be ranked equally with bodily integrity or practical reason is seen as ‘incredible’ by others (Gough, 2003: 14). Interestingly, there seems to be no greater agreement in the substantive content of these lists within a particular perspective (need based, capability, flourishing) than across these different perspectives.
Appendix III: Objective list or substantive good accounts

Substantive good accounts encompass a range of measures from different perspectives: needs based, perfectionist or flourishing, capabilities and lists. Although many of the measures drawn from substantive good accounts share similarities, there is also considerable variation in what is taken to be the elements of well-being. Conflicting versions that simultaneously claim universal applicability suggests a lack of complete consensus on attributes which are inherently good, and consequently measures which are potentially influenced by the beliefs and values of those generating the scales.

A single index of well-being, although generated by some of these measures, is in conflict with the multi-dimensionality and non-commensurability of attributes within these accounts. Where an index is generated, the relative weight given to different attributes lacks theoretical justification. Such indexes are not well suited to aggregation over a specified duration of time, and lifetime well-being is best represented through the addition of length of life as an additional, non-commensurable attribute, hence trade-offs between quality and length of life are not possible.

III.2 How closely do substantive good measures meet the criteria for an ideal measure?

III.2.1 Introduction

This next section will consider the performance of substantive good measures against the criteria set out in Chapter Four.

III.2.2 Appropriate conception of well-being for policy

Prudential value only

Non-prudential concerns may be considered to be part of leading a ‘good life’ from a substantive good perspective (Sumner, 1995: 773). However, they do not need to be included in the judgement of what is good for the individual except to the extent that they contribute directly to how well life goes for the individual concerned.

Where human capacity incorporates virtue this seems to incorporate values which extend beyond prudential values. Aristotelian versions of eudaemonia are states which are worth having in one’s life in part because of their moral virtue, and as such give a normative statement of how one ought to live, rather than a statement of when life is best for the individual.
Appendix III: Objective list or substantive good accounts

A difficulty arises because, as noted above, most non-prudential concerns also have prudential aspects, even virtue may be an entirely prudential concern. For example, the ability to, and act of caring for others could be seen as part of what it is to live a truly human existence and necessary for life to go well for the individual.

The lists and measures described under the substantive good account do not contain any aspects which pertain to non-prudential values, therefore seem to reflect how well life goes for the individual.

Inter-temporally and interpersonally comparable

Substantive good accounts of well-being aim to be universally relevant to human experience, as such they should be inter-personally comparable. The important question is whether any single individual or group of individuals from a particular period in time and location is able to identify the attributes which are universally relevant. Lists are open to objections that they reflect the social, economic and cultural values of the authors. Concerns focus on how those determining the constituents of well-being could avoid biases arising from their own perspective which will have arisen from being embedded in social conditions. If the list is a representation of the culture and time in which it is written this will limit its applicability for cross-cultural comparisons, and comparisons over time. Nussbaum’s list, for example, has been criticised on the basis of American values applied to developing countries (Stewart, 2001).

In Rhetoric, Aristotle includes ‘many children’, ‘large stature’ and ‘athletic powers’ within his list of the constituent parts of happiness (Aristotle, 1954, Book 1, Chapter 5, paragraph 3). These appear very historically and culturally specific. For example, in a recent survey in Thailand 69% of respondents considered having few children as necessary or very necessary to well-being (Camfield et al., 2007).

Lists may be more relevant to some groups within a population than others undermining the ability to use such lists to make inter-personal comparisons. For example, Nettle (2005) argues that the ‘lower’ hedonic pleasures are principally associated with the poorly educated and the ‘higher’ eudemonic pleasures with the better educated. Nevertheless, if a list is widely agreed upon within a group of people interpersonal comparisons within that group would seem reasonable. However, if a list of attributes can be agreed upon which are interpersonally comparable, even across cultures, it will also be necessary to measure these attributes in a way which maintains the cultural neutrality.

The UK’s Indices of Deprivation are a good example of how considerations differ across fairly small geographical areas, with separate indices used for England, Wales, Scotland and Northern
Appendix III: Objective list or substantive good accounts

Ireland. Where lists focus upon resources rather than attainments, comparisons are particularly problematic. An attribute such as the presence of central heating is culturally and climate specific suggesting that such lists are unlikely to translate into cross country comparisons.

Many substantive good accounts aim to represent essential human concerns, therefore should in theory be interpersonally comparable. However, in practice someone must decide upon the list of attributes which are held to be well-being and it will always be difficult to separate this process from cultural, time, location specific influences. When basic needs involve resources and inputs to well-being these may become even more culturally specific. Using psychological needs may remove the list from culturally specific resource inputs, however, the importance of psychological needs, such as autonomy, may still vary according to culture.

**III.2.3 Validity**

*Content validity*

Nussbaum describes Aristotle’s eudaemonic tradition as seeing happiness as “something like flourishing human living … inclusive of all that has intrinsic value, and complete, meaning lacking in nothing that would make it richer or better” (Nussbaum, 2005: 171). By definition, therefore, a flourishing account of well-being should be complete.

Substantive good accounts of well-being have been criticised on the basis that feelings and subjective judgements matter to how well an individual’s life goes, in addition to objective circumstances. This can be addressed by the inclusion of subjective assessments as additional substantive goods (although this solution opens the measure to the problems of adaptation). Whether a substantive list of constituents of well-being has content validity will always be open to debate, since people differ on the contents of what should be included upon any list. For example, religious belief may be taken an essential attribute for some and unimportant for others.

*Convergent validity*

The extent to which a substantive good measure will converge with other indicators of well-being will depend upon the extent to which that indicator is included within any substantive measure. However, there is little empirical work to drawn upon.

Where a measure is reflecting only capability and opportunities for outcomes these may not necessarily correlate with improved outcomes. For example, the opportunity for healthy diet
may exist without this being achieved. Hence there may be a lack of convergence between capability and other objective well-being outcomes.

*Predictive validity*

Substantive good accounts have the potential to predict outcomes of importance, in part because they may incorporate characteristics which predict future states. For example, academic performance may be included in a substantive good list for adolescents, not because it is thought to be a current component of well-being rather because it is thought to predict future well-being. The predictive validity of the few measures that offer a single well-being index is an area where more research is required.

**III.2.4 Empirically useful**

*Reliable with low and unbiased measurement error*

Due to the lack of single index scales, there is minimal evidence upon which to draw information on reliability.

*Cardinal*

Substantive good accounts do not attempt to give a single figure for well-being. Consequently, they do not suggest any method for weighing up improvements in one area compared with improvements in another. Whilst there are obvious problems in assuming commensurability between different attributes (between say autonomy and personal relationships), it is also not helpful for their use in public policy to insist they can not be commensurable at any level. Unless the account includes some degree of tradability it will remain an incomplete theory of well-being for policy purposes. Where arbitrary trade-offs are used within a measure without a clear theoretical framework there are many theoretical levels of well-being which one individual could be thought to have.

Within substantive good accounts more of a constituent of well-being, or more opportunity for attaining that constituent, is assumed to result in a higher level of well-being. For example, it is often assumed that the presence of more autonomy will make a persons life better, regardless of their tastes and preferences. However, at certain points more of an attribute may arguably not be an improvement in an individual’s well-being. For example, Griffin (1986) uses the example of more autonomy creating anxiety in some individuals. When using opportunity, such as more choice of attaining more autonomy, it is appealing to argue that if it does not make an individual’s life better then it would not be chosen. However, the very presence of choice may
not be neutral to one’s experience and more choice may not be better (Schwartz, 2004). A possible solution is to present a capability for a simple life which may be undermined by the presence of too much choice. However, if this is negatively correlated with increased opportunity in other areas, and if it is not possible to make trade-offs between different capabilities it will be hard to ever know whether an individual’s well-being has increased or decreased. More of a constituent of well-being should enhance well-being, however, it may be that the relationships between some of the constituents of well-being and well-being are non-linear, possibly becoming negative at some levels. For example, the first seven years of education may be strongly related to how well an individual’s life goes for them, but further education beyond 18 years of education may no longer be beneficial and may be detrimental to success in other domains of life. This would imply that separate consideration would be required for each constituent of well-being.

Cardinal measures of the constituents of well-being (health, education, freedom, personal relationships etc.) can be created for convenience of measurement. One option may be to use time as a tool to achieve cardinality, such as days an individual spends in particular states (e.g. being disease free), however, such convenience measures will lack theoretical foundations.

As seen in Chapter Two some measurement instruments drawn from a substantive good perspective generate a single index. This implies a latent variable of well-being, which exists as an amalgamation of the constituents of well-being. Whilst the scales (e.g. Ryff’s PWB) may appear cardinal, there is no clear rationale for how components within the scale should be weighted, or that treating the scale as cardinal is theoretically meaningful. Generating a single index from a substantive list account of well-being will always be open to debate and potential criticism. However, the benefits of such an index may outweigh theoretical criticism.

Appropriate sensitivity

Substantive good measures of well-being can be sensitive to policy relevant changes in circumstances by incorporating these circumstances into the measure of well-being. This shows the clear trade-off between validity of the measure, and sensitivity to changes in a factor of interest. Measures such as the PWB scale have a wide potential range, with 120 questions each with 6 options (shortened versions have 18 questions each with 6 options), this may therefore be sensitive although there is likely to be a trade-off between sensitivity and length of instrument.

Distributions of evaluative measures are negatively skewed (e.g. SLWS), however, some of the more eudaemonic measures, such as the OHS have a more normal distribution (Peterson et al., 2005), suggesting greater sensitivity at higher end of the scale.
Appendix III: Objective list or substantive good accounts

Practical and efficient to collect

Measures derived from a substantive good account perspective have usually focused on social well-being and taken advantage of existing social level data. Measurement burden depends upon the extent of this list, how compatible it is with existing data sources, and whether it includes subjective and well as objective constituents. Measures derived from a flourishing perspective, focusing on psychological flourishing, involve self-report and moderately quick surveys.

Public/Political acceptability

Substantive good accounts have the potential to reflect agreement from within the society over what is of value. However, that presupposes such agreement is possible. Where opinions differ, particularly if they differ systematically according to a characteristic of the population, such as education, universal support will be unlikely. Consequently, the policy relevance of scales which incorporate hedonic pleasure or eudaemonic pleasure depends, at least in part, on the acceptability of these underlying values for policy.

A substantive list of the constituents of well-being may easily overstep the remit of concern as judged by those in favour of liberal government, where items on the list include such attributes as personal achievement, and virtue. Whilst liberals may acknowledge that such attributes are part of well-being, they may not agree that this conception of well-being has a role within government policy.

Where opportunities are used, rather than outcomes, this has the advantage of offering a balance between individual and collective responsibilities. Opportunities may be present but it will be up to the individual to make the most of those opportunities. Whilst this may have certain appeal, as discussed in Chapter Two, there is no recognised means to measure true opportunities as oppose to functionings.

III.2.5 Conclusion

As noted in Chapter Five the most important of all criteria must be validity. Without validity a measure that is successful on all other criteria may be lead to misleading policy.

It is likely that there exists a version of a substantive good account which is a valid measure of well-being, but its not clear how we decide upon which version that is. Neither is it clear what weights should be given to different attributes upon any list. The theoretical potential for validity of the account has not yet led to any measures in which we have empirical evidence of validity.