Essays on predictability of equity and bond risk premia

Andrea Carnelli
February 2014

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of Imperial College London
Declaration of originality

I herewith certify that this thesis constitutes my own work and that all material, which is not my own work, has been properly acknowledged.

Andrea Carnelli
Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Abstract

The purpose of this thesis is to investigate the evidence of return predictability in equity and treasury markets. The first topic investigates the evidence on return predictability from an economic perspective. We use a simple model that incorporates a time varying investment opportunity set into a mean-variance portfolio maximization framework, and derive the optimal capital allocation weights for: (i) a naive strategy based on average realized returns; and (ii) a class of strategies that condition on dividend-price signals. While our data supports in-sample evidence of return predictability, the out-of-sample returns of the naive strategy are higher than those of all conditional portfolio specifications based on a certainty equivalent metric and portfolio turnover. These results suggest that dividend-price predictability offers no economic value to investors. The second topic studies the link between short and long-run risk premia of equity claims. We extract short-term risk premia from contemporaneous information on short-term futures and cash equity markets under the assumption of no arbitrage. Predictability regressions reveal that short-term risk premia capture different information from long-run risk premia. Counter to the intuition that a high price of risk commands high returns, high short-run risk premia on dividend claims predict low returns on the index. While inconsistent with models featuring either habit persistence or long-run risk, the results may be reconciled with some models of uncertainty aversion.

The third topic is concerned with monetary policy sources of bond risk premia. Expected monetary policy shocks are extracted from a panel of inflation, GDP growth, and federal funds rates forecasts and using a Taylor rule specification as an identifying assumption. Expected monetary policy shocks are found to be statistically significant predictors of excess returns on bonds, even after controlling for levels and conditional volatilities of macroeconomic activity. The findings are consistent with a long run risk economy where contractionary monetary policy action increases GDP uncertainty.
Acknowledgments

First and foremost, I am indebted to my supervisor, Professor Andrea Buraschi, Chair in Finance at Imperial College Business School. His inspiration, guidance, and continuous encouragement throughout these years have instilled in me the curiosity for thought-provoking questions, and patiently taught me the tools required to answer them. This thesis is the product of hundreds of hours of discussion with him.

My gratitude also goes to Paul Whelan for the countless days and nights spent trying to further our understanding of this intriguing “science”.

Last but least, I thank James Grant, Simon Schillebeeckx, Andres Reibel, Farouk Jivraj and all other fellow PhD candidates at Imperial College for having made my time here an incredible experience.
to Dorian, Birgit, Claudio, and Sergio
# Contents

Declaration of originality 1

Copyright 2

Abstract 3

Acknowledgments 4

1 Introduction and literature review 11

2 The economic value of predictability 16
   2.1 Introduction ................................................. 16
   2.1.1 Related Literature ........................................... 17
   2.2 Dividend yields and portfolio management .................. 19
      2.2.1 Capital allocation models ............................... 19
      2.2.2 Performance evaluation ................................... 22
   2.3 Empirical results ............................................. 23
      2.3.1 Data .................................................. 23
      2.3.2 Learning from dividend yields .......................... 23
      2.3.3 The economic value of dividend yields .................... 28
   2.4 Conclusion .................................................. 35

3 Short versus long run risk premia 37
   3.1 Introduction ................................................ 37
   3.2 A Motivating Example ....................................... 40
4.7.2 Money non-neutrality and habit: an example

5 Conclusion

130
List of Tables

2.1 DP ratio predictability ........................................ 25
2.2 DP yield predictability ......................................... 26
2.3 Portfolio return statistics ...................................... 30
2.4 Economic value .................................................. 31

3.1 Information in stpd ............................................. 52
3.2 Predictability of S&P500 returns: STRP ...................... 55
3.3 Predictability of S&P500 returns: LTPD ...................... 57
3.4 Predictability of S&P500 returns: bivariate regressions .... 58
3.5 Bond predictability: univariate regressions ................ 60
3.6 Bond predictability: bivariate regressions ................. 61

4.1 Summary statistics of bond excess returns .................. 85
4.2 Taylor rule specifications ...................................... 90
4.3 Summary statistics of expected monetary policy shocks .... 92
4.4 Spanning ......................................................... 95
4.5 Bond return predictability: $u^e$ ............................ 104
4.6 Bond return predictability: $u^e$ and $g$ ..................... 106
4.7 PathShocks and QE path shocks ............................. 108
4.8 Monetary policy shocks and equity returns ................ 111
4.9 Macroeconomic uncertainty ................................... 116
4.10 Excess returns, habit and LRR .............................. 119
4.11 Monetary policy shocks and surplus ......................... 120
4.12 Monetary policy shocks and uncertainty ................... 121
4.13 Monetary policy shocks versus uncertainty ............... 129
List of Figures

2.1 Market weights .................................................. 32
2.2 Cumulative returns ............................................. 33
2.3 Performance attribution ....................................... 34

3.1 Dividend prices .................................................. 47
3.2 Dividend prices: 1 year maturity ............................ 48
3.3 Price-to-dividend ratio of dividend claim (1 year maturity) ........................................ 51
3.4 Risk premia ....................................................... 53
3.5 Admissible correlations ......................................... 65
3.6 Loadings of expected returns on \( x_t \) .......................... 66

4.1 Consensus federal funds forecasts .............................. 79
4.2 Consensus macro forecasts ..................................... 80
4.3 The performance of BCFFS forecasts ......................... 82
4.4 Macroeconomic activity ........................................ 84
4.5 Bond excess returns ............................................ 86
4.6 Expected monetary policy shocks ......................... 93
4.7 Procyclicality of PathShock ................................. 94
4.8 Monetary policy shocks and yield curve information .... 97
4.9 Expected monetary policy shocks and excess returns .... 99
4.10 Constituents of mps ............................................ 110
4.11 Habit ............................................................. 115
4.12 GDP expectations .............................................. 117
4.13 Inflation expectations ......................................... 118
Chapter 1

Introduction and literature review

The quest of asset pricing consists in determining suitable risk premia for financial claims with uncertain payoffs. A large body of empirical evidence suggests that expected returns vary over time and that can be predicted in the time series of bond and equity returns. Predictability is an intriguing feature of the data to both financial economists—who read it as an empirical restriction for equilibrium models of preferences and technologies—and finance practitioners—who can exploit it for tactical portfolio allocation. This thesis aims to shed light on various aspects of predictability in both equity and bond markets by studying the empirical evidence, economic value, and economic underpinnings of time varying risk premia.

Are equity returns predictable? The first answer to this question, provided as early as the beginning of the 20th century, was “no”. Louis Bachelier, a French mathematician, conjectured that asset prices follow a random walk in his 1900 PhD thesis “The Theory of Speculation”. Bachelier’s dissertation, along with some initial empirical studies, were published in 1964 in an anthology edited by Paul Cootner. The efficient market hypothesis (EMH) emerged as a theory and, in 1970, Fama published a review of both the theory and evidence of the EMH (Malkiel & Fama (1970)). Following Fama’s influential article, a vast empirical literature has been focusing on whether asset prices follow a random walk.

In the mid-1980s, however, an increasingly large number of empirical works were finding that, in contrast to the random walk view, stock returns are predictable by financial ratios and in particular by the dividend yield and the price-earning ratio (Fama & French

The debate on the predictability of bond returns has followed a similar path. Until the 1980s, the prominent theory of the term structure of interest rates was the Expectation Hypothesis, according to which the holding period returns on Treasury securities, while possibly maturity-specific, are constant over time. Many variables, however, have been found to have predictive power for bond excess returns: the current term structure (Fama & Bliss (1987), Campbell & Shiller (1991), Cochrane & Piazzesi (2005)); macroeconomic variables (Ludvigson & Ng (2009)); disagreement (Burasi & Whelan (2012)); realized jumps (Wright & Zhou (2009)) and volatility (Mueller et al. (2012)); liquidity (Fontaine & Garcia (2012)). Taken as a whole, this large body of literature suggests that the data strongly reject Expectation Hypothesis.

Was the paradigm of market efficiency dead? Indeed, the initial economic interpretation of the strong evidence of predictability was that these tests were rejecting the efficient market hypothesis. Fama (1991), however, suggested that, far from rejecting the main paradigm, the evidence was simply rejecting the assumption that expected returns are constant: time-variation in expected returns can be consistent with market efficiency. Over the past thirty years, several asset pricing theories have proposed alternative explanations for the time variation in conditional expected returns. The most influential streams of this literature include: (i) “habit models” in which the effective level of risk aversion is time-varying and countercyclical (see Campbell & Cochrane (1999)); (ii) “long-run risk models” in which small but persistent shocks to consumption growth are priced (see Bansal & Yaron (2004)); (iii) “rare disaster models” (see Gabaix (2012)); (iv) “heterogeneous beliefs models” (see Jerome
& Shashidhar (1994), and Buraschi & Jiltsov (2006)). In all these models, time-varying expected returns are an equilibrium outcome and predictability is consistent with market efficiency.

Recent work suggests that the finding that risk premia vary over time is a spurious implication of biases in regression models or anomalous subsamples (see, for instance, Goetzmann & Jorion (1993), Nelson & Kim (1993), Valkanov (2003), Ang & Bekaert (2007), Welch & Goyal (2008)). It is not clear, however, whether these claims are damning for the empirical relevance of predictability. Cochrane (2008), for instance, argues that classic regression-based tests fail to recognize that a test of return predictability implies a joint null hypothesis, under which returns are not predictable and dividend growth is predictable at the same time. As a consequence, to the extent that dividend-price ratios fail to forecast dividend growth, little evidence of return predictability does not mechanically imply that expected returns are constant over time. Chapter 2 attempts to overcome this impasse by asking whether predictability would have been economically valuable to an investor faced with real-time capital allocation choices. The use of real-time data dismisses concerns about look-ahead biases; economic value added, on the other hand, represents an intuitive metric to assess the extent of predictability, instead of the classic measures based on the economic and statistical significance of the slope coefficients in predictive regressions. We start by reviewing the in-sample empirical evidence. Next, we assess the economic value of predictability from the perspective of an investor who shifts her exposure to equity depending on real time signals. We construct market timing strategies by using the portfolio allocation model developed by Brandt & Santa-Clara (2006), which is an extension of the Markowitz paradigm that accommodates time varying expected returns. We consider a "naive" strategy, which allocates capital on the basis of rolling averages of realized returns, and strategies that condition on price-dividend signals. Finally, we evaluate the out-of-sample performance of the strategies by using the three economic metrics suggested by DeMiguel et al. (2009): Sharpe ratios of realized returns, certainty equivalents, and portfolio turnover. Contrary to what suggested by the in-sample results, which are particularly strong, especially for the 1947-1990 period, we find that a portfolio manager would have not been able to, and willing to, exploit dividend price signals in real-time. With the exception of the 50s, 70s, and 80s, an investor would have been better off by following the naive strategy throughout the 20th century up until
today.

Stocks can be viewed as portfolios of zero-coupon dividend securities, i.e. claims that pay an uncertain dividend at future dates. This representation parallels that of coupon-bearing debt securities, which can be thought of as portfolios of zero-coupon bonds. Just like bond returns tend to be correlated across the maturity spectrum, it is natural to expect some degree of comovement in the risk premia associated with dividend claims of different maturity. Building upon this intuition, Chapter 3 extracts the expected return on 1-year dividend claims and assesses its forecasting power for index excess returns. Following Binsbergen et al. (2012b) and Binsbergen et al. (2012a), we obtain synthetic dividend prices from strategies that are contemporaneously long the index and short the corresponding futures contract. Next, we construct short term expected returns by projecting the return of dividend claims on their lagged price-to-dividend ratios. Finally, we investigate the forecasting power of short term expected returns for the holding period returns of the index. Contrary to the intuition that dividend risk premia are correlated across maturities, we present strong statistical evidence that high expected returns on dividend claims with short maturity forecast low excess returns on the index. These results, which suggest that equity risk premia follow complex dynamics, are even more puzzling when evaluated in light of the leading asset pricing models such as long-run-risk (Bansal & Yaron (2004)) or habit (Campbell & Cochrane (1999)) economies.

Chapter 4 investigates predictability in the context of fixed income markets. In particular, it asks whether monetary policy is a source of time variation in the price of risk. The question arises naturally because bond yields are risk neutral expectations of future short rates, which are set by the central bank: monetary policy may thus have an impact on yields via physical expectations of future short rates and risk adjustments. An empirical proxy of monetary policy path shocks is obtained from the residuals of Taylor rule regressions estimated over a survey of Federal Funds, GDP and inflation forecasts. Predictability regressions show that monetary policy shocks forecast bond excess returns in an economically important and statistically significant way. The exposure to monetary policy shocks is also priced in the cross-section of equity returns, commanding a risk premium of approximately 4%. Finally, the chapter asks whether the empirical evidence is consistent with a quantity of risk or price of risk channel, and finds that monetary policy path shocks increase the conditional volatility
of GDP growth, a source of time varying risk premia in long run risk economies.

Finally, chapter Chapter 5 presents some concluding remarks.
Chapter 2

The economic value of predictability

2.1 Introduction

Are stock returns predictable? If so, do dividend-price signals really add economic value to an investor who needs to make capital allocation choices? This chapter accomplishes two goals. First, we survey the literature on equity predictability and provide an update on empirical evidence by using sample data until 2012; the presence of Post-Crisis information allows us to investigate the statistical robustness of previous results. Second, we study the economic value of dividend-price predictability by considering the optimal portfolio choice decision of an agent who uses dividend-price ratios to model conditional expected returns. If predictability can be indeed exploited in real time, information encoded in predictive state variables is of considerable interest to practitioners who can develop market-timing portfolio strategies to enhance profits.

The capital allocation model we adopt is based on the work of Brandt & Santa-Clara (2006), and accommodates a time varying investment opportunity set. The economic intuition is simple: the optimal fraction of the portfolio that is allocated to the risky asset is increasing in expected returns, so that portfolio choice is sensitive to the information content of predictors such as dividend yields. We compare the out-of-sample performance of portfolios based on this conditioning information to that of benchmark “naive” portfolios that allocate capital based on a moving average of realized returns. If dividend yields actually

\footnote{The work in this chapter appears in Buraschi & Carnelli (2013a).}
predict returns in a way that is economically valuable for an investor, a strategy based on
dividend-yield signals should be preferable to one that relies on the assumption that returns
are i.i.d. We follow DeMiguel et al. (2009) and use three metrics to gauge economic value:
Sharpe ratios of realized returns, certainty equivalents, and portfolio turnover.

The results can be summarized as follows. First, we confirm the results of previous studies
by finding evidence of return predictability in the 1947-1990 sample. Both the economic
and statistical significance of the slope coefficients are increasing in maturity, and \( R^2 \) are
high as 60\% for holding period horizons of 5 years. Dividend growth predictability, on the
other hand, is almost non-existent. These results, which are robust to the choice of the
predictor (dividend-price ratios or yields), confirm the widespread notion that time variation
in dividend-price ratios captures time variation in expected returns. We also find, however,
that these results are highly sample specific: when the sample is extended to the period 1926-
2012, the level of return predictability is cut by half. Second, we assess the economic value
of predictability for an investor who makes real-time capital allocation decisions based on
dividend-price signals. We find that some signals generate marginal value to investors when
evaluated on the basis of Sharpe ratios. From a certainty equivalent and portfolio turnover
perspective, however, the naive portfolio always beats the more sophisticated conditional
specifications. We present a simple performance attribution graphic diagnostic that shows
that treating ratios as trading signals outperformed naive strategies only during the 50s and
70s/80s. An investor who had used dividend ratios as signals for capital allocation would
have consistently underperformed a naive strategy from the early 90s until today.

The rest of the chapter is organized as follows. After a survey of related literature in
the next subsection, Section 2 describes the portfolio model and the metrics of economic
value-added we adopt. Next, section 3 presents the data and the empirical results. Section
4 concludes.

2.1.1 Related Literature

This chapter is related to two streams of the asset pricing literature: predictability and
optimal portfolio choice.
As discussed in the introduction of the thesis, there is a large body of evidence that documents the predictability of equity returns. Today, the emerging consensus is that "Most financial economists appear to have accepted that aggregate returns do contain an important predictable component." [Campbell (2000) p. 1523]. At the same time, several scholars have suggested that the evidence on the predictability of stock returns based on in-sample regressions may be spurious. In the context of dividend ratios, for example, Goetzmann & Jorion (1993) and Ang & Bekaert (2007) have criticized the conclusions based on these specifications and several scholars have highlighted the low in-sample power of many of these tests (see also Nelson & Kim (1993), and Valkanov (2003)). Welch & Goyal (2008) use data up to 2006 to study the predictive variables proposed by the empirical literature. They find that the evidence is sample specific and some of the statistical significance depends on using data up to the 1973-1975 Oil Shock period. Their results motivate the importance to study the economic value of predictability in the context of optimal portfolio formation. Estimation error and spurious predictive results could negatively affect the out-of-sample portfolio performance of an investor seeking to use these models for market timing. The contribution of this chapter is to provide rigorous empirical evidence on this issue.

We also draw from a second stream of the asset pricing literature that studies optimal portfolio choice when the investment opportunity set is time varying. In principle, return predictability can be exploited in the construction of optimal portfolios. In practice, computing optimal dynamic portfolios that consistently exploit the return predictability is no easy task. The reason is that closed-form solutions are available only for rather special cases. Most approaches proposed in the literature use different type of numerical methods. Michael J. Brennan & Lagnado (1997), Barberis (2000), and Lynch (2001) use discrete-state approximations to numerically solve for the optimal portfolio choice problem of a long horizon investor when returns are predictable. Campbell & Viceira (1999), Campbell & Viceira (2001), and Campbell & Viceira (2002) use analytical approximations in the neighborhoods of known exact solutions in the context of an infinite horizons portfolio choice problem. Unfortunately, the numerical complexity of these methods is such that most practitioners have eventually gone back to use either the simple and well-known Markovitz allocation or naive portfolios. Recently, Brandt & Santa-Clara (2006) have proposed a methodology that allows to exploit predictability in a time-consistent manner within the context of portfolio
optimization. Their idea is an adaptation of the conditionally managed portfolios described in Hansen & Richard (1987): given a state variable that is presumed to forecast returns, they form a portfolio that invests in a set of primitive assets for an amount that is proportional to the level of the conditioning variable. The procedure consists of choosing portfolio weights as simple linear functions of the predictive variables. Thus, the optimal portfolio solution is a simple maximization of the agent’s utility function with respect to the parameters of this linear function. This approach allows us to address the economic question that we want to investigate: does return predictability add value to portfolio selection or, rather, do estimation and misspecification errors reduce the out-of-sample performance relative to naive portfolios? In judging predictability based on the economic value that accrues to investors, we follow a relatively young but flourishing strand of literature. DeMiguel et al. (2009) study the out-of-sample performance of the sample-based mean-variance model relative to a naive equally-weighted portfolio. The authors find that the naive portfolio outperforms the mean-variance model and all its extensions aimed at reducing estimation error. The interpretation of these results is that, in the samples that are available, the costs associated with estimation error outweigh the benefits from efficient allocation. In this chapter, we incorporate a time varying opportunity set into the portfolio choice problem and apply the approach of DeMiguel et al. (2009) to study the economic value of dividend-price ratios for capital allocation strategies. In recent work, Della Corte et al. (2008) and Thornton & Valente (2012) use a similar framework in the context of fixed income markets to assess the economic value of bond return predictability.

2.2 Dividend yields and portfolio management

2.2.1 Capital allocation models

This section describes the reference model for capital allocation. Since the Markowitz model is inconsistent with time-varying expected returns, we use a one-asset version of the single-period conditional portfolio problem studied by Brandt & Santa-Clara (2006). In a nutshell, this approach reduces the optimal solution of a dynamic strategy problem to a static choice of managed portfolios by assuming that the portfolio weights are proportional to the level of the conditioning variable. This highly stylized model can be thought of as describing
the problem of an investment manager who seeks to maximize next-period returns while exploiting predictability signals at the same time.

Let \( R_{t+1} = (P_{t+1} + D_{t+1}) / P_t \) denote the total return on the market asset between \( t \) and \( t+1 \), and let \( x_t \) denote the portfolio weight of the risky asset. The investor chooses the optimal portfolio weight by solving a standard quadratic maximization problem:

\[
\max_{x_t} E_t \left[ x_t R_{t+1} - \frac{\gamma}{2} x_t^2 \tilde{R}_{t+1}^2 \right],
\]

(2.1)

where \( \gamma \) is a risk aversion parameter; by formulating the problem in terms of total returns, we are implicitly assuming that the remainder of the portfolio's value is held as cash with zero return. The Markowitz paradigm is a special case of (2.1): returns are assumed to be i.i.d., so that conditional expectations can be replaced with their unconditional counterparts and the problem can be solved easily by looking at first order conditions. The complication arises when returns are not i.i.d., the case we consider. We follow Brandt & Santa-Clara (2006) and assume that portfolio weights are proportional to the state variable \( Z_t \) capturing time variation in expected returns:

\[
x_t = \theta Z_t.
\]

By substituting the parametric assumption above into \( x_t \) and introducing the notation \( \tilde{R}_{t+1} = Z_t R_{t+1} \), the maximization problem can be re-written as:

\[
\max_{x_t} E_t \left[ \theta \tilde{R}_{t+1} - \frac{\gamma}{2} \theta^2 \tilde{R}_{t+1}^2 \right].
\]

As Brandt & Santa-Clara (2006) highlight, \( \tilde{R}_{t+1} \) can be interpreted as the return on a managed portfolio, which invests in the market asset an amount that is proportional to the value of the state variable. Since the same \( \theta \) maximizes the conditional expected utility at all dates \( t \), it also maximizes the unconditional expected utility:

\[
\max_{x_t} E \left[ \theta \tilde{R}_{t+1} - \frac{\gamma}{2} \theta^2 \tilde{R}_{t+1}^2 \right],
\]

thus reducing the dynamic strategy problem to a simply static problem. The solution to this maximization problem follows easily from the F.O.C.:

\[
\theta = \frac{1}{\gamma} E \left[ \tilde{R}_{t+1}^2 \right]^{-1} E \left[ \tilde{R}_{t+1} \right].
\]

(2.2)
Since our aim is to study the economic value-added of predictability to investors, we follow DeMiguel et al. (2009) and estimate (2.2) recursively over rolling windows of fixed size in order to avoid look-ahead biases. Fixing the size of the rolling window to \( M \), the portfolio weight at time \( t \), \( x_t = \theta Z_t \), is obtained by estimating \( \theta \) via the sample counterpart of (2.2):

\[
\hat{\theta}_{t, t} = \frac{1}{\gamma} \left[ \sum_{s=0}^{M} \hat{R}_{t-s}^2 \right]^{-1} \left[ \sum_{s=0}^{M} \hat{R}_{t-s} \right],
\]

so that the time \( t \) weight based on signal \( Z \) is given by \( \hat{x}_{t, t} = \hat{\theta}_{t, t} Z_t \). We consider two investment strategies. The first strategy, which acts as a benchmark, is a “naive” strategy that selects the exposure to the market based on observed realized excess returns; this case can be seen as a degenerate managed portfolio with \( Z = 1 \). The second strategy uses dividend-price ratios as signals. We consider three signals: the dividend-price ratio \( (DP_t) \), the earnings-price ratio \( (EP_t) \), and the (inverse of the) cyclically-adjusted price-earnings ratio \( (1/CAPE_t) \). We set the risk aversion parameter to 3; this choice ensures that, in our sample, the portfolio share invested in the market is always between 0 and 1, so that the strategies we consider are self-financing.

How exactly do these capital allocation strategies relate to the classic regressions of realized returns on dividend-price ratios? Consider the expression (2.2); treating the denominator as a constant of proportionality, and expanding the numerator \( E \left[ \hat{R}_{t+1} \right] = E \left[ Z_t R_{t+1} \right] \), we obtain:

\[
\theta \propto E \left[ Z_t R_{t+1} \right]
\]

\[
\propto Cov \left[ R_{t+1}, Z_t \right] + E \left[ R_{t+1} \right] E \left[ Z_t \right]
\]

\[
\propto const. + b Var \left[ Z_t \right],
\]

where \( b = Cov \left[ R_{t+1}, Z_t \right] / Var \left[ Z_t \right] \) is the slope coefficient of a projection of realized returns on dividend-price ratios. This expression highlights that \( \theta \) is an affine transformation of the slope coefficient from traditional predictability regressions: consistent with the intuition, the higher the economic significance of predictability in the sample \( (b) \), the more sensitive is the optimal market allocation to time variation in the dividend-price ratio \( (\theta) \).
2.2.2 Performance evaluation

This section describes the metrics employed to evaluate the economic value of capital allocation strategies. The out-of-sample returns of the capital allocation strategies are defined as realized market returns scaled by the lagged weight, plus uninvested cash:

\[ R_{t+1, Z} = \hat{x}_{t, Z} R_{t+1} + (1 - \hat{x}_{t, Z}). \]

Let \( \hat{\mu}_Z \) and \( \hat{\sigma}_Z \) denote the sample mean and standard deviation of out-of-sample returns \( R_{t+1, Z} \):

\[
\hat{\mu}_Z = \frac{1}{T - M} \sum_{s=M+1}^{T} R_{s, Z}
\]

\[
\hat{\sigma}_Z = \sqrt{\frac{1}{T - M} \sum_{s=M+1}^{T} \left( R_{s, Z} - \frac{1}{T - M} \sum_{s=M+1}^{T} R_{s, Z} \right)^2}
\]

We follow DeMiguel et al. (2009) and use three metrics to measure the economic value of our capital allocation strategies. First, we construct Sharpe ratios of realized out-of-sample returns:

\[ \hat{S}R_Z = \frac{\hat{\mu}_Z}{\hat{\sigma}_Z}. \]

Next, we compute the certainty equivalent:

\[ \hat{C}E_Z = \hat{\mu}_Z - \frac{\gamma}{2} \hat{\sigma}_Z^2. \]

Finally, we construct a turnover metric:

\[ \hat{T}\hat{O}_Z = \frac{1}{T - M} \sum_{s=M+1}^{T} |\hat{x}_{s+1, Z} - \hat{x}_{s, Z}|, \]

where \( \hat{x}_{s+1, Z} \) is the market weight before rebalancing at time \( s + 1 \).

\[ \text{The wealth} \ W \text{ of the investor grows according to} \ W_{t+1} = W_t R_{t+1, Z}. \text{ Before rebalancing, the time} \ t + 1 \text{ market weight is given by} \ x_{t+1} = \frac{W_t R_{t+1}}{W_{t+1}} = x_t \frac{R_{t+1}}{R_{t+1, Z}}. \]
2.3 Empirical results

2.3.1 Data

All price, dividend, and earnings data refer to the S&P500 index and are taken from Robert Shiller’s website. Given the long span of the sample (1871-2012), we use real (CPI-deflated) variables to ensure that results are not due to inflation effects. Since index dividends and earnings tend to feature high seasonality at monthly and quarterly frequency, we work with annual (end of year) data. Dividends (earnings) are 12-month moving sums of dividends paid on (earnings generated by) the S&P 500 index; this aggregation procedure implicitly assumes that interim dividends are kept as cash rather than being re-invested in an interest bearing account or in the S&P500.

2.3.2 Learning from dividend yields

In order to gain intuition about the economic interpretation of classical predictive regressions, we first review the decomposition of Campbell & Shiller (1988a). Let lower-case letters indicate log-variables, so that \( r_{t+1} = \log(R_{t+1}) \), and define the (log) dividend-price ratio as \( dp_t \equiv d_t - p_t \). Campbell & Shiller (1988a) show that returns can be written as

\[
    r_{t+1} = \alpha + \Delta d_{t+1} - \psi dp_{t+1} + dp_t
\]

where \( \alpha \equiv \log(1 + \exp(-dp^*)) + \psi dp^* \), \( \psi \equiv \frac{\exp(-dp^*)}{1+\exp(-dp^*)} \), and \( dp^* \) is the long-run average of \( dp_t \). Equation 2.4 is a differential equation which can be solved either forward or backward; iterating forward and imposing a transversality condition (no bubbles) we obtain:

\[
    dp_t - dp^* = E_t \sum_{s=1}^{\infty} \psi^{s-1} [(r_{t+s} - r^*) - (\Delta d_{t+s} - d^*)].
\]

This equation has been studied in a variety of contexts. Two implication emerge. First, if log returns \( r_t \) and dividend growth \( \Delta d_t \) are stationary, then log dividend yields \( dp_t \) are stationary. Second, deviations of the dividend yield from its long run mean \( dp_t - dp^* > 0 \) imply that either future returns \( r_{t+s} \) or dividend growth \( \Delta d_{t+s} \) will exceed their long run mean. Using annual data for the sample period 1927-2004, Cochrane (2008) argues that since dividend yields only weakly predict future dividend growth at the aggregate level, time variation in
dividend yields should mechanically help to explain the time-variation in conditional expected returns.

We review the empirical evidence on the informational content of dividend-price ratios by running two classic predictive regressions:

\[ R_{t+h} = a + b \left( \frac{D_t}{P_t} \right) + \epsilon_{t+1} \]
\[ D_{t+h}/D_t = a + b \left( \frac{D_t}{P_t} \right) + \epsilon_{t+1}. \]

We consider horizons \( h \) from 1 to 5 years. Since Goyal & Welch (2003) find that dividend-price ratios \( (D_t/P_t) \) and dividend yields \( (D_t/P_{t-1}) \) contain different information, and that the empirical evidence in favour of predictability is strongest for the sample up until the 1990s, we report results for both predictors and two samples: 1947-1990, and 1926-2011. Tables 2.1 and 2.2 contain the results, for dividend price ratios and yields, respectively.

When we use dividend-price ratios as predicting variable, we find strong evidence that returns can be forecasted in the 1947-1990 sample period. At a one year horizon, the t-statistics of the slope coefficient is 3.51 (we use Newey & West (1987) corrected HAC consistent standard errors) and the \( R^2 \) is equal to 16%. As we increase the holding period horizon, the degree of predictability increases and at an horizon of five years, the t-statistic is 6.15 with an \( R^2 \) equal to 60%. At the same time, the nature of this predictability is clearly not related to the predictability of dividend growth. Indeed, at any horizon, the slope coefficient of a regression of future dividend growth onto current dividend-price ratio is not statistically different from zero. It is interesting to observe, however, that the result is substantially weakened as we extend the sample period to include both the period before WWII and the more recent period after 1990 until 2011. In the extended sample period, while the slope coefficient is generally significant, the predictability is lost for the holding period horizon of 1 year. Moreover the \( R^2 \) is substantially reduced: it is equal to 9% at a two year holding period horizon and reaches 17% at a five year horizon. Interestingly, the slope coefficients of future dividend growth predictive regressions switch sign and turn negative. A negative slope coefficient is consistent with economic theory: a drop in prices, thus an increase in the dividend-price ratio, should forecast a drop in future dividends.

When we consider dividend yields as the forecasting factor, the previous results are confirmed. In general, we find that the dividend yield is a slightly weaker predictor. In the
Table 2.1: DP ratio predictability
The table reports the output of regressions of cum-dividend returns (left panel) or dividend growth (right panel) on a constant and on dividend-price ratios. Horizons ($h$) range from 1 to 5 years. T-statistics, reported below the point estimates, use Newey & West (1987) HAC-consistent standard errors ($h$ lags). The top (bottom) panel reports the results for the 1947-1990 (1926-2011) sample.

<table>
<thead>
<tr>
<th>DP RATIO</th>
<th>Returns</th>
<th>Dividend growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947-1990</td>
<td>$h$</td>
<td>$a$</td>
</tr>
<tr>
<td>1</td>
<td>0.84</td>
<td>5.80</td>
</tr>
<tr>
<td></td>
<td>11.21</td>
<td>3.51</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
<td>11.62</td>
</tr>
<tr>
<td></td>
<td>4.79</td>
<td>3.44</td>
</tr>
<tr>
<td>3</td>
<td>0.63</td>
<td>14.97</td>
</tr>
<tr>
<td></td>
<td>3.58</td>
<td>3.94</td>
</tr>
<tr>
<td>4</td>
<td>0.46</td>
<td>21.16</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>5.42</td>
</tr>
<tr>
<td>5</td>
<td>0.17</td>
<td>30.09</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>6.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1926-2011</th>
<th>$h$</th>
<th>$a$</th>
<th>$b$</th>
<th>$R^2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>2.15</td>
<td>3.54%</td>
<td>1.10</td>
<td>-1.94</td>
<td>9.35%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.51</td>
<td>1.61</td>
<td>29.14</td>
<td>-1.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.95</td>
<td>5.29</td>
<td>9.17%</td>
<td>1.12</td>
<td>-2.17</td>
<td>4.64%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.68</td>
<td>2.62</td>
<td>18.56</td>
<td>-1.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.96</td>
<td>6.93</td>
<td>11.02%</td>
<td>1.13</td>
<td>-1.86</td>
<td>2.45%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.53</td>
<td>2.63</td>
<td>14.93</td>
<td>-0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>10.41</td>
<td>16.34%</td>
<td>1.13</td>
<td>-1.67</td>
<td>1.70%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.33</td>
<td>3.06</td>
<td>15.32</td>
<td>-0.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.90</td>
<td>12.35</td>
<td>17.66%</td>
<td>1.14</td>
<td>-1.47</td>
<td>1.28%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.25</td>
<td>2.76</td>
<td>15.05</td>
<td>-0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.2: DP yield predictability
The table reports the output of regressions of cum-dividend returns (left panel) or dividend growth (right panel) on a constant and on dividend-price yields. Horizons (h) range from 1 to 5 years. T-statistics, reported below the point estimates, use Newey & West (1987) HAC-consistent standard errors (h lags). The top (bottom) panel reports the results for the 1947-1990 (1926-2011) sample.

<table>
<thead>
<tr>
<th>DP YIELD</th>
<th>Returns</th>
<th>Dividend growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947-1990</td>
<td>h</td>
<td>a</td>
</tr>
<tr>
<td>1</td>
<td>0.89</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>11.29</td>
<td>2.75</td>
</tr>
<tr>
<td>2</td>
<td>0.87</td>
<td>7.16</td>
</tr>
<tr>
<td></td>
<td>7.83</td>
<td>3.02</td>
</tr>
<tr>
<td>3</td>
<td>0.76</td>
<td>11.61</td>
</tr>
<tr>
<td></td>
<td>5.49</td>
<td>4.23</td>
</tr>
<tr>
<td>4</td>
<td>0.52</td>
<td>19.03</td>
</tr>
<tr>
<td></td>
<td>2.85</td>
<td>5.77</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
<td>26.49</td>
</tr>
<tr>
<td></td>
<td>1.30</td>
<td>6.93</td>
</tr>
<tr>
<td>1925-2011</td>
<td>h</td>
<td>a</td>
</tr>
<tr>
<td>1</td>
<td>0.97</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>18.59</td>
<td>2.29</td>
</tr>
<tr>
<td>2</td>
<td>0.97</td>
<td>4.85</td>
</tr>
<tr>
<td></td>
<td>10.72</td>
<td>2.47</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>7.28</td>
</tr>
<tr>
<td></td>
<td>7.24</td>
<td>2.68</td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>10.16</td>
</tr>
<tr>
<td></td>
<td>4.99</td>
<td>2.71</td>
</tr>
<tr>
<td>5</td>
<td>0.91</td>
<td>11.82</td>
</tr>
<tr>
<td></td>
<td>3.85</td>
<td>2.30</td>
</tr>
</tbody>
</table>

26
1947-1990 sample period the slope coefficient is strongly significant and the $R^2$ ranges between 12% and 59%. The statistical significance is extremely strong for holding period returns of 5 years. Also the statistical significance of dividend yields is substantially weakened in the extended sample 1926—2011: the highest $R^2$, at the 5 years horizon, does not exceed 13%. Interestingly, predictive regressions for future dividend growth produce slope coefficients which are positive (thus with the wrong sign), but the null of no significance is never rejected. In the extended sample period, the failure of dividend yields to forecast future fundamentals is very noticeable with $R^2$ never exceeding 1% at any holding period. This evidence reinforces the argument that time variation in the dividend-price ratio captures time variation in the price of risk rather than the dynamics of fundamentals.

The somewhat lack of robustness of predictive regression across sub-samples has induced scholars to explore the out-of-sample performance of predictors. Welch & Goyal (2008) run out-of-sample regressions in which expected future returns at time $t + 1$ are based only on information available at time $t$. They find that in some subsamples a naive forecast obtained from the sample mean of returns up to time $t$ can do better than a predictive regression that uses dividend-price ratio as a conditioning variable. Cochrane (2008) argues, however, that ‘Out-of-sample $R^2$ is not a test; it is not a statistic that somehow gives us better power to distinguish alternatives than conventional full-sample hypothesis tests.’ (page 1538). Indeed, the lower out-of-sample performance of the dividend-price ratio cannot be used as a test of the null hypothesis that returns are i.i.d. As the return decomposition 2.4 shows, the null hypothesis of i.i.d. returns has additional implications that need to be jointly tested. Cochrane (2008) argues that, among these other hypotheses, the most important to rule out return predictability is a large predictability of future dividend-growth and a small “long-run” return predictability. Thus, on the basis of the results presented in Tables 2.1 and 2.2, one can hardly argue against Cochrane’s conclusion that ‘there is in fact less than a 5% chance that our data or something more extreme is generated by a coherent world with unpredictable returns.’ (page 1538).

The out-of-sample results reported by Welch & Goyal (2008) serve, however, as a clear warning on the practical use of dividend-price ratios in portfolio management. The forecasting variable is persistent and dividends are known to be the outcome of corporate decisions that tend to smooth dividend distributions. These results raise an important question: If
expected returns are time-varying and dividend yields are an important state variable, what is the economic value of predictability in the context of optimal portfolio choice? We address this question in the next section.

2.3.3 The economic value of dividend yields

We consider four alternative strategies to construct the optimal portfolios. The first portfolio is a benchmark naive portfolio that does not use conditioning information to predict future returns. In this case the market allocation is based on a simple average of realized returns (NAIVE), the other three portfolios use cash-flow ratio signals as conditioning information. The first signal is the dividend-price ratio (DP), the second signal is the earning-price ratio (EP), and the third is (inverse of) the cyclically adjusted price-to-earnings ratio (1/CAPE). Optimal portfolio weights are estimated recursively via rolling windows of $M$ annual observations, with $M$ being either 30, 60, or 90 years. In all cases, the information used to construct a portfolio is strictly based on data available up to that moment, thus ruling out any potential look-ahead bias.

Table 2.3 summarizes the sample statistics of 1-year out-of-sample portfolio returns of the four capital allocation strategies. Using a rolling window of 30 years and holding the portfolio for 1 year, we find that the average real return is 2.26% for the 1/CAPE portfolio, followed by the DP and EP portfolios with average real returns equal to 2.21% and 2.08%, respectively. Interestingly, however, all the three portfolios produce lower mean returns than the NAIVE portfolio. The result is even sharper when using median returns. While the NAIVE portfolio produces a median real return of 2.77%, the DP portfolio produces a median real return of 1.90%. Some interesting results emerge also from higher order moments. The standard deviation of the the NAIVE portfolio is clearly higher than the three portfolios based on conditioning information. This is especially true for $M = 60$ and 90. This suggests that while conditioning information is of limited use to forecast the first moment of expected returns, it seems to be valuable to reduce the dispersion of the distribution of expected returns. This effect is quite striking when we look at the "skewness" and "minimum" returns. At horizons of $M = 60$ and 90, the skewness of the NAIVE portfolio is highly negative, while the skewness of the portfolios based on conditioning information is positive, for $M = 60$, and close to zero, for $M = 90$. The economic impact of the negative skewness is evident when we look at the
minimum returns, which are -11.62% for the NAIVE portfolio at $M = 60$, versus minimum returns of -7.24% for the DP portfolio. The result is confirmed when we consider $M = 90$. These findings motivate the need to compare the four portfolios on the basis of economic value metrics.

Table 2.4 summarizes economic value statistics of the 1-year out-of-sample returns of the four portfolios. We consider three statistics: (i) Sharpe ratio (SR); (ii) certainty equivalent (CE); and (iii) portfolio turnover (TO). Using rolling windows of $M = 30$, we find that the Sharpe ratio of the NAIVE portfolio is roughly equivalent to that of the three dynamic portfolios. For $M = 60$ and 90, however, the Sharpe ratios of the dynamic portfolios are generally larger than the NAIVE one. For $M = 90$, in particular, the Sharpe ratio of the DP portfolio is 0.45 versus 0.39 for the NAIVE portfolio. When we compare the four strategies based on their certainty equivalent (thus accounting for the agent’s risk aversion), we find that the NAIVE strategy has the largest CE value. We also find that the NAIVE strategy has also another important property: it implies the lowest portfolio turnover.

The dynamics of portfolio weights are summarized by Figure 2.1. We find that while the NAIVE strategy, by construction, implies a portfolio weight which is very persistent, the allocation to the risky asset implied by the use of conditioning variable is highly time-varying. We also find that, in general, the signals provided by DP, PE, and 1/CAPE ratios are indeed quite similar. Some noticeable exceptions emerge. At the end of the 2007-2008 Crisis, the EP ratio was suggesting a sharp reduction in exposure to risky assets. The signal coming from the DP and 1/CAPE ratios, on the other hand, was suggesting an aggressive increase in exposure to the equity market. Aside from this noticeable case, however, the three signals feature a strong positive correlation, with the EP ratio implying the most aggressive dynamic reallocation. Interesting examples are at the end of WWII and at the end of the second Oil Shock, when the PE ratio implied a bullish exposure to equity markets; in both cases, the optimal allocation to risky assets was far greater than what implied by the DP ratio.

Figure 2.2 shows the difference in cumulative performance of the four strategies. As of 2012, the NAIVE portfolio would have produced larger gains than the three dynamic portfolios. The EP portfolio trails the NAIVE portfolio closely and, for an extended period (up until 1997) produces the largest cumulative returns. Once the compounding effect is taken into account, the difference in the final value of a 1 dollar invested in the 20s is quite
Table 2.3: Portfolio return statistics

The Table contains the sample statistics of 1-year out-of-sample portfolio returns of four capital allocation strategies. The first column contains the results for a naive strategy that chooses the market allocation based on a simple average of realized returns; the second, third and fourth column use cash-flow ratio signals as conditioning information. Optimal portfolio weights are estimated recursively via rolling windows of \( M \) annual observations; the top, medium, and bottom panel report results for windows of size 30, 60, and 90 years, respectively.

<table>
<thead>
<tr>
<th></th>
<th>NAIVE</th>
<th>DP</th>
<th>EP</th>
<th>1/CAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M = 30 )</td>
<td>Mean 2.35%</td>
<td>2.21%</td>
<td>2.08%</td>
<td>2.26%</td>
</tr>
<tr>
<td></td>
<td>Median 2.77%</td>
<td>1.90%</td>
<td>1.74%</td>
<td>1.79%</td>
</tr>
<tr>
<td></td>
<td>Stdev 5.74%</td>
<td>5.22%</td>
<td>5.53%</td>
<td>5.49%</td>
</tr>
<tr>
<td></td>
<td>Skew -0.24</td>
<td>0.22</td>
<td>-0.67</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Kurtosis 2.70</td>
<td>3.89</td>
<td>6.40</td>
<td>4.84</td>
</tr>
<tr>
<td></td>
<td>Min -11.66%</td>
<td>-12.88%</td>
<td>-23.79%</td>
<td>-13.98%</td>
</tr>
<tr>
<td></td>
<td>Max 15.62%</td>
<td>19.07%</td>
<td>16.74%</td>
<td>21.45%</td>
</tr>
<tr>
<td></td>
<td>Nobs 111</td>
<td>111</td>
<td>111</td>
<td>101</td>
</tr>
<tr>
<td>( M = 60 )</td>
<td>Mean 2.55%</td>
<td>2.18%</td>
<td>2.33%</td>
<td>2.13%</td>
</tr>
<tr>
<td></td>
<td>Median 3.42%</td>
<td>1.74%</td>
<td>1.79%</td>
<td>1.84%</td>
</tr>
<tr>
<td></td>
<td>Stdev 5.51%</td>
<td>4.47%</td>
<td>4.64%</td>
<td>3.75%</td>
</tr>
<tr>
<td></td>
<td>Skew -0.24</td>
<td>0.87</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Kurtosis 2.93</td>
<td>5.27</td>
<td>3.05</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>Min -11.62%</td>
<td>-7.24%</td>
<td>-9.15%</td>
<td>-7.30%</td>
</tr>
<tr>
<td></td>
<td>Max 15.60%</td>
<td>19.95%</td>
<td>15.44%</td>
<td>12.05%</td>
</tr>
<tr>
<td></td>
<td>Nobs 81</td>
<td>81</td>
<td>81</td>
<td>71</td>
</tr>
<tr>
<td>( M = 90 )</td>
<td>Mean 1.91%</td>
<td>1.29%</td>
<td>1.55%</td>
<td>1.66%</td>
</tr>
<tr>
<td></td>
<td>Median 3.08%</td>
<td>1.45%</td>
<td>1.85%</td>
<td>1.49%</td>
</tr>
<tr>
<td></td>
<td>Stdev 4.92%</td>
<td>2.90%</td>
<td>3.88%</td>
<td>3.74%</td>
</tr>
<tr>
<td></td>
<td>Skew -0.73</td>
<td>-0.14</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>Kurtosis 3.03</td>
<td>3.16</td>
<td>4.31</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>Min -11.53%</td>
<td>-6.56%</td>
<td>-9.95%</td>
<td>-8.14%</td>
</tr>
<tr>
<td></td>
<td>Max 10.36%</td>
<td>8.48%</td>
<td>13.12%</td>
<td>11.35%</td>
</tr>
<tr>
<td></td>
<td>Nobs 51</td>
<td>51</td>
<td>51</td>
<td>41</td>
</tr>
</tbody>
</table>

30
Table 2.4: Economic value

The Table contains economic value statistics of 1-year out-of-sample portfolio returns of four capital allocation strategies. The statistics include: Sharpe ratio (SR), certainty equivalent (CE), and portfolio turnover (TO). The first column contains the results for a naive strategy that chooses the market allocation based on a simple average of realized returns; the second, third and fourth column use cash-flow ratio signals as conditioning information. Optimal portfolio weights are estimated recursively via rolling windows of $M$ annual observations; the top, medium, and bottom panel report results for windows of size 30, 60, and 90 years, respectively.

<table>
<thead>
<tr>
<th></th>
<th>NAIVE</th>
<th>DP</th>
<th>EP</th>
<th>1/CAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M = 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>0.41</td>
<td>0.42</td>
<td>0.38</td>
<td>0.41</td>
</tr>
<tr>
<td>CE</td>
<td>1.85%</td>
<td>1.80%</td>
<td>1.62%</td>
<td>1.81%</td>
</tr>
<tr>
<td>TO</td>
<td>3.50%</td>
<td>7.47%</td>
<td>8.11%</td>
<td>7.08%</td>
</tr>
<tr>
<td>$M = 60$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>0.46</td>
<td>0.49</td>
<td>0.50</td>
<td>0.57</td>
</tr>
<tr>
<td>CE</td>
<td>2.10%</td>
<td>1.89%</td>
<td>2.01%</td>
<td>1.91%</td>
</tr>
<tr>
<td>TO</td>
<td>3.40%</td>
<td>6.08%</td>
<td>7.19%</td>
<td>5.24%</td>
</tr>
<tr>
<td>$M = 90$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>0.39</td>
<td>0.45</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>CE</td>
<td>1.55%</td>
<td>1.17%</td>
<td>1.33%</td>
<td>1.45%</td>
</tr>
<tr>
<td>TO</td>
<td>3.13%</td>
<td>4.39%</td>
<td>5.77%</td>
<td>5.32%</td>
</tr>
</tbody>
</table>
Figure 2.1: Market weights

The Figure plots the estimated market allocation weights $\hat{x}_{t,Z}$ for four capital allocation strategies: a naive strategy that chooses the market allocation based on a simple average of realized returns, and three strategies that use cash-flow ratio signals as conditioning information. Optimal portfolio weights are estimated recursively via rolling windows of 60 annual observations.

striking. While the real value of the initial investment increased about 4 times using the DP and 1/CAPE strategy, the NAIVE strategy produced an increase in real terms of 5.3 times the initial capital invested.
Figure 2.2: Cumulative returns
The Figure plots the cumulative returns of four capital allocation strategies: a naive strategy that chooses the market allocation based on a simple average of realized returns, and three strategies that use cash-flow ratio signals as conditioning information. Optimal portfolio weights are estimated recursively via rolling windows of 60 annual observations.

In order to investigate in greater detail the relative performance of the three dynamic strategies, we multiply the difference in market allocation of conditional strategies relative to the naive portfolio by the market return: \((x_{t,Z} - x_{t,NAIVE}) R_{t+1}\); positive values of this quantity indicate that the conditional strategy has allocated more (less) capital to the market
Figure 2.3: Performance attribution

Each line plots the spread between the market allocation weights of a conditional and a naive strategy times the market return one period later: \((\hat{x}_{t,Z} - \hat{x}_{t,NAIVE})R_{t+1}\). Values above zero indicate that the conditional strategy has outperformed the naive strategy in a given period. The naive strategy chooses the market allocation based on a simple average of realized returns, while the conditional strategies use cash-flow ratio signals as conditioning information. Optimal portfolio weights are estimated recursively via rolling windows of 60 annual observations.
in bullish (bearish) markets. Figure 2.3 plots \((x_{t,Z} - x_{t,NAIVE})R_{t+1}\) for \(M = 60\). We find that dynamic strategies have outperformed the NAIVE strategy in only two periods: immediately after WWII, between 1948 and 1952, and in the 1970s during the two Oil Shocks; in the rest of the sample, the NAIVE strategy has produced higher returns.

These results suggest that while a consensus has emerged in finance that the dividend-price ratio is a powerful predictive variable which helps to explain the time-variation in expected excess returns, a real time portfolio strategy based on dividend-price ratio would have not outperformed a strategy based on the assumption that market returns are i.i.d.

### 2.4 Conclusion

The empirical literature of the past thirty years has built a strong body of evidence in support of the notion that equity returns are predictable; advances on the theoretical front, on the other hand, have provided equilibrium foundations to justify the idea that returns are not i.i.d. Recent studies, however, have cast doubts on these findings on statistical grounds: subsample analysis shows great variation in the level of predictability, and the out-of-sample forecast power of dividend-price ratios is disappointing.

This chapter tackles the issue of predictability from a capital allocation perspective: we ask whether a portfolio manager would have been able and willing to exploit dividend-price ratios to enhance performance. We find that the answer to this question is, alas, "no". A sophisticated investor who had consistently relied on dividend-price signals over the past century would now find himself lagging behind the naive investor. When examined in the light of the economic value of out-of-sample returns, in-sample predictability is a mirage that can be traced back to 50s and 70s/90s.

The results are based on the assumption of a 1-year investment horizon. This assumption might bias the rejection of the null of predictability in two ways. First, we know from in-sample regressions that predictability is strongest for long horizons, so that an increase in the holding period returns of the investor might increase the economic value of dividend-price signals for portfolio performance. Second, by assuming a myopic horizon, we are ignoring the impact that hedging demand may have on optimal weights. We however believe that our model, while highly stylized, is effective in describing the real-world problem of active
portfolio managers whose performance is evaluated over short horizons: for this class of investors, strategies based on dividend-price ratios are likely to prove disappointing.
Chapter 3

Short versus long run risk premia

3.1 Introduction

This chapter investigates the relationship between risk compensation on equity claims of different maturities: we construct a measure of expected excess returns on claims to one year dividends (the "short term" asset) and assess its forecasting power for index excess returns (the "long term" asset). Understanding the properties of risk premia is one of the most important yet challenging tasks in asset pricing. One of the challenges is related to the fact that, in general, it is difficult to identify risk premia demanded by investors: they cannot be directly observed when the timing and magnitudes of dividends are unknown. A stock index pays uncertain dividends at uncertain times over a life of uncertain duration. Investors form expectations about future cash flows and use a set of discount rates to determine the value. All information is then aggregated by a tatonnement and the econometrician can only observe a time series of stock prices. This loss of information is a binding constraint for our understanding of asset prices; its limits are laid bare when contrasted to our understanding of discount rates in fixed income markets. In the fixed income literature, the term structure of interest rates is an observable entity with two important roles. First, it encodes information about the dynamics of the stochastic discount factor (yields are risk adjusted expectations of future short rates). Second, it provides a host of restrictions that allow to better discriminate among alternative models. If equity discount rates were observable, we could learn about

\[1\]The work in this chapter appears in Buraschi & Carnelli (2013b).
their dynamics and generate additional restrictions that asset pricing models have to satisfy.

How can we identify risk premia on short term assets? Following the pioneering work of Binsbergen et al. (2012b) and Binsbergen et al. (2012a), we address this challenge by using observations on dividend prices. Dividend prices can be obtained by imposing a simple no-arbitrage pricing restriction on cash and derivatives markets. Unlike stock prices, the prices of dividends carry information about future expected dividend and risk premia over fixed horizons. By projecting excess returns of dividend claims on their price to dividend ratio, we are able to identify the level of compensation required by investors to bear equity risk of finite duration. Since the dividend prices we focus on are claims on one year dividends, we label the risk premia we obtain as short term (STRP). Next, we investigate the empirical properties of STRP and their relationship to LTRP, the risk premia earned by holding the equity index (the long term asset). We explore this relationship indirectly by running predictive regressions of realized excess returns on lagged STRP. Since realized excess returns reveal changes in equity discount rates at all maturities, these regressions allow us to uncover the dynamic link between short- and long-term discount rates. Finally, we ask to what extent the dynamics of risk premia on short term dividends help to explain index returns and which asset pricing model can be consistent with some of its basic properties.

Our main empirical exercise is a regression of excess returns on one year risk premia: can theory say something about the sign and magnitude of the slope coefficient? The problem can be best understood by means of an analogy to fixed income securities. We can think of an equity index as a long maturity bond (say, 30 years) that pays coupons (i.e. dividends), and we can think of STRP as the yield on a short maturity (say, 1 year) zero coupon bond. The issue we address is akin to the question: can one year yields predict one-year holding period returns on a 30 years maturity bond? A rigorous answer to the question involves specifying a term structure model that: i) derives the price of a bond as a function of state variables, and ii) determines risk premia by modeling the covariance between shocks to bond prices and shocks to the stochastic discount factor. To the extent that the yield on the 1 year bond reveals the state variable, it is also possible to form expectations about the 1 year holding period return of the 30 year maturity bond. An equity index is like a coupon-bearing bond, except that it has infinite maturity and cash flow risk. Just like a coupon-bearing bond can be thought of as a portfolio of zero coupons, an equity index is a portfolio of “zero
coupon" dividend strips, i.e. claims to dividends payable at future dates. The term structure model of dividend strips we adopt is the one developed in Lettau & Wachter (2007). Lettau & Wachter (2007) show that short term risk premia (our regressor) identify the price of risk up to a scaling factor, and that expected returns on dividend strips with longer maturities are positive functions of the price of risk. Since the return on the index is a weighted average of returns on dividend strips of all maturities, the model provides a tight link between the left and right hand side variables of our regression, and motivates its predictive power.

Our results can be summarized as follows. We find that short term expected returns predict realized returns over short horizons in an economically and statistically significant way. The predictive power of STRP is over and above that of price-to-dividend ratios: STRP detect high frequency predictability components in stock returns that price-to-dividend ratios do not capture. Even more interesting, the sign of the slope coefficient is negative: higher STRP are followed by lower index excess returns. The results are difficult to be reconciled with leading asset pricing models, at least in their traditional specifications. These models generate risk premia dynamics and term structures of expected equity returns that are too restrictive. In long run risk models, for instance, there is little room for predictability by short term risk premia. Growth and uncertainty shocks (the priced risks in the economy of Bansal & Yaron (2004)) affect long term cash flows and have little impact on short term cash flows: the link between expected returns on high- and low-duration claims is weak. In consumption-habit models, on the other hand, expected returns on short-term claims can predict index returns; the sign of the slope coefficient, however, must be positive, as the price of risk is the only state variable driving the term structure of expected returns.

We identify the two ingredients that are necessary for a model to be able to explain this type of predictability: shocks to dividends must be priced, and there must be a multivariate state vector driving the term structure of risk premia. Short-term dividend claims appear to be substitute assets to short term bonds. The result is consistent with an economy in which, during periods of high uncertainty, uncertainty-averse agents have a preference for short duration assets (short-term dividend claims) to the expense of long-term duration assets (equity shares). Short-term dividends have an advantage with respect to longer-term cash-flow: they are less uncertain. Indeed, when risk premia rise and agents flee to quality, we see the emergence of a preference for short-term duration assets. Depending on the extent to
which agents reduce the cash-flow duration of their exposure, this may give rise to a negative slope coefficient in long-term predictive regressions based on short-term risk premia.

This work is especially related to the study of equity yields in Binsbergen et al. (2012a). The dataset of Binsbergen et al. (2012a) consists of dividend futures prices that two investment banks use for pricing and mark-to-market purposes. The authors construct equity yields as the ratio of the future price of dividend futures to the current dividend. Since time variation in equity yields reflects time variation in dividend growth expectations and/or risk premia, the authors use a VAR to separate dividend growth from risk premia components, and find that equity yields are good forecasters of dividend and consumption growth, while risk premia are counter cyclical. Our analysis is different from Binsbergen et al. (2012a) in three respects. First, we use data from index futures, as opposed to dividend futures, due to the longer sample period and larger liquidity. Implied dividend prices are obtained using the cost of carry formula. Second, instead of looking at the predictability of dividend growth, we focus on the predictability of returns. Finally, our analysis focuses on the predictability of index returns conditional on short term risk premia thus allowing us to study the properties of the dynamics and time dependence of the term structure of equity risk premia.

The rest of the chapter is organized as follows. Section 2 motivates the predictability regressions as an implication of the model of Lettau & Wachter (2007). Section 3 describes the data and the construction of STRP. Section 4 presents the empirical results, while section 5 discusses the economic interpretation. Section 6 concludes.

3.2 A Motivating Example

An investor wishing to invest in a broad equity index is faced with two options. The first, and most obvious strategy, is to long the index. Since stocks can be viewed as portfolios of dividend claims (see, for instance, Lettau & Wachter (2007) and Binsbergen et al. (2012b)), the holding period return earned is equal to a weighted average of the holding period returns earned on dividend claims of all maturities. Dividend claims do not generate interim cash flows: the holding period return is entirely due to changes in value, which must converge to realized dividends at maturity. A second alternative is to invest in a single dividend claim.
This chapter focuses on the relationship between the risk premia earned by investing in one-year dividend claims ("short term risk premia", STRP) and risk premia earned by holding a long position in the index ("long term risk premia", LTRP).

How are STRP and LTRP related? As a motivation for the regressions that follow, we use the term structure model of dividend prices developed by Lettau & Wachter (2007). The logarithm of dividend growth $\Delta d_{t+1} = \log(D_{t+1}/D_t)$ and its conditional mean $z_{t+1}$ are assumed to follow

$$
\Delta d_{t+1} = g + z_t + \sigma_d \epsilon_{t+1},
$$

$$
z_{t+1} = \phi_z z_t + \sigma_z \epsilon_{t+1},
$$

where $g$ is the unconditional growth rate of log dividend growth, and $\phi_z$ captures the degree of persistence of shocks to the time-varying component of expected dividend growth. The price of risk is driven by a state variable $x_t$ that evolves according to

$$
x_{t+1} = (1 - \phi_x) \bar{x} + \phi_x x_t + \sigma_x \epsilon_{t+1},
$$

where $\bar{x}$ is the unconditional mean of the price of risk, while $\phi_x$ captures the persistence of innovations to the price of risk. In all expressions, $\epsilon_{t+1}$ is a $3 \times 1$ vector of independent standard normal shocks, so that the "shock loadings" $\sigma_d$, $\sigma_z$, $\sigma_x$ are $1 \times 3$ row vectors. Note that the conditional volatility of each variable is given by the norm of the vector of shock loadings, while the conditional covariance between any two variables is given by the inner product of the respective vectors of shock loadings. Since only dividend risk is assumed to be priced, and the (log) risk free rate $r^f = \log R^f$ is assumed to be constant, the stochastic discount factor equals

$$
M_{t+1} = \exp\left\{ -r^f - \frac{1}{2} \sigma_d^2 - x_t \frac{\sigma_d}{||\sigma_d||} \epsilon_{t+1} \right\}.
$$

Let $P_{t,n}$ be the price of a dividend claim that pays the aggregate dividend $n$ periods from now, and let $R_{n,t+1} = P_{t+1,n-1}/P_{t,n}$ denote the return on holding the dividend claim for one period. Note that by no arbitrage, it must be that $P_{t,0} = D_t$. Lettau & Wachter (2007) show that under these assumptions both the price-to-dividend ratio and the expected excess return on dividend claims are exponential-affine functions of the state variables:

$$
\log P_{t,n} = \log D_t + A(n) + B_x(n)x_t + B_z(n)z_t
$$
\[
\log E_t \left[ R_{n,t+1}/R^f \right] = \left( \sigma_d + B_x(n-1)\sigma_x + B_z(n-1)\sigma_z \right) \frac{\sigma_d'}{||\sigma_d||} \tilde{x}_t, \tag{3.3}
\]
where the values of the coefficients \(A(n), B_x(n), \) and \(B_z(n)\) are given by the boundary condition for \(n = 0\)
\[
A(0) = B_x(0) = B_z(0) = 0,
\]
and, for \(n > 0\), by the recursions
\[
A(n) = A(n-1) - r^f + g + B_x(n-1)(1 - \phi_z)\tilde{x} + \frac{1}{2} V_{n-1} V_{n-1}'
\]
\[
B_z(n) = \frac{1 - \phi_z^2}{1 - \phi_z},
\]
\[
B_x(n) = B_x(n-1) \left( \phi_x - \sigma_x \frac{\sigma_d'}{||\sigma_d||} \right) - \left( \sigma_d + B_z(n-1)\sigma_z \right) \frac{\sigma_d'}{||\sigma_d||}.
\]
Since an index can be thought of as a portfolio of dividend claims, its value today is equal to the sum of the prices of the dividend strips
\[
P_t = \sum_{n=1}^{\infty} P_{t,n},
\]
and the expected excess return it generates is a weighted average of dividend risk premia across the range of maturities
\[
\log E_t \left[ R_{t+1}/R^f \right] = \log \left( \sum_{n=s}^{\infty} w_n \exp \left\{ \left( \sigma_d + B_x(n-1)\sigma_x + B_z(n-1)\sigma_z \right) \frac{\sigma_d'}{||\sigma_d||} x_t \right\} \right), \tag{3.4}
\]
where \(w_n = P_{n,t}/P_t\) and \(s = 1 (s = 2)\) for cum-dividend (ex-dividend) returns.

Equations (3.3) and (3.4) allow us to gain insight into the relationship between STRP and LTRP. STRP are the empirical counterpart to (3.3) for \(n = 1\)
\[
\log E_t \left[ R_{1,t+1}/R^f \right] = ||\sigma_d|| x_t. \tag{3.5}
\]
Equation (3.5) says that risk premia backed out of short term dividend expectations and prices are a constant multiple of the latent factor \(x_t\). Since the expected return on the index is also a function of \(x_t\), its derivative with respect to \(||\sigma_d|| x_t\) is different from zero
\[
\beta = \frac{\partial \log E_t \left[ R_{n,t+1}/R^f \right]}{\partial ||\sigma_d|| x_t} \tag{3.6}
\]
Defining \( \text{const.} = \log E_t \left[ \frac{R_{t+1}}{R_t} \right] - \beta_1 \sigma_d \| x_t \), and letting \( \{ u_t \} \) denote a white noise process orthogonal to \( x_t \), one can see that the model imposes a restriction on the coefficients of the regression
\[
\log(R_{t+1}/R_t) = \text{const.} + \beta \sigma_d \| x_t + u_{t+1},
\]
Equation (3.7) is the main object of interest of our empirical exercise.

The previous relationship offers an interesting set of empirical predictions that can be used to learn about the properties of the term structure of equity risk premia. First, in the RE (rational expectations) exponentially affine model described above, the only risk involved in buying a claim to one-year forward dividends is dividend volatility. Thus, short term risk premia are proportional to the latent factor that drives the price of risk. Moreover, since the price of risk is a state variable for the entire term structure of dividend strips, variation in short term risk premia captures variation in realized returns. Second, risk premia implied by P/D ratios and those implied by claims to short term dividends should provide the same information. A rejection of these testable implications, on the other hand, may suggest that the term structure of equity risk premia is affected by multiple factors operating at different horizons.

### 3.3 Data and construction of STRP

We construct STRP from dividend prices implied by S&P500 futures. The sample comprises quarterly observations from June 1992 to March 2012. When constructing dividend prices, we choose one-year horizons to avoid biases arising from the seasonality of dividends. A side effect of this approach is that while S&P500 futures start trading as far back as 1982, our sample only starts in June 1992 because the fourth futures contract (i.e. the contract with one year to maturity) is not continually available for trade until that date. The use of futures data motivates the choice of sampling frequency: since contract maturities are in the March quarterly cycle, quarterly observations dispense us with the need to obtain constant maturity contracts, which would introduce interpolation errors. An alternative to backing out dividend prices indirectly from index futures via the cost of carry would be to obtain direct observations from dividend swaps, as in Binsbergen et al. (2012a). Dividend swaps
have the advantage of being traded for a large range of maturities; at the same time, they are OTC contracts with counter-party risk and have gained in popularity quite some time after index futures. Also, dividend prices implied by dividend swaps are potentially subject to price pressure owing to institutional hedging. The following subsections describe the construction of the variables in detail.

### 3.3.1 Dividend prices

Under the assumption of no arbitrage, the price of a dividend is implicit in the spread between spot and (discounted) futures prices. More formally, consider a dividend paying financial asset with time \( t \) price \( S_t \), and a forward contract written on the asset with delivery \( n \) periods in the future. Define \( P_{t,n} \) as the value that makes the cost of carry formula true:

\[
F_{t,n} = [S_{t+n} - P_{t,n}] e^{ny_t^{(n)}},
\]

where \( y_t^{(n)} \) is a (continuously compounded) risk-free rate appropriate to discount \( t + n \) cash flows back to \( t \). The cost-of-carry formula can be re-written in terms of \( P_{t,n} \) as

\[
P_{t,n} = S_t - F_{t,n} e^{-ny_t^{(n)}}. \tag{3.8}
\]

By no arbitrage \( P_{t,n} \) must be the price that the market is willing to pay for the dividend stream \( \{D_{t+j}\}_{j=0}^n \) payable by the underlying between \( t \) and \( t + n \):

\[
P_{t,n} = E_t \left[ \sum_{j=0}^{n} M_{t+j} D_{t+j} \right],
\]

where \( M_t \) is the time \( t \) stochastic discount factor. Equation (3.8) thus provides a means to infer the price of dividend claims from spot and forward market data.

In order to construct dividend prices using (3.8) we need the time series of spot prices \( S_t \), futures prices \( P_{t,n} \), and interest rates \( y_t^{(n)} \) that are appropriate to discount time \( t + n \) cash flows to \( t \). S&P500 futures trade on the Chicago Mercantile Exchange (CME). Available

---

\(^3\)Investment banks typically sell derivatives whose payoffs are ex-dividend; hedging these derivatives requires a position in the underlying, which gives rise to a long exposure to dividend risk. If dividend risk is too high, the investment bank might have the incentive to sell dividend derivatives at particularly attractive prices.
maturities vary within the sample. When S&P 500 futures started trading in 1982, available maturities included the first 4 months in the March cycle. Ever since, the number of available contracts has increased progressively. All open positions are settled in cash at close of the last day of trading (3:15 p.m. on Thursday prior to third Friday of the contract month) to the Special Opening Quotation (SOQ). The SOQ is calculated using normal index procedures except that the prices used are the actual opening values of the constituents. Since securities might start trading minutes or hours after the opening bell of the NYSE, the SOQ can be significantly different than the published opening value of the index. The SOQ is used because it represents actionable values and thus ensures convergence of cash and futures markets.

Both futures and spot data are taken from Thomson Datastream. We combine underlying and futures prices with interest rates data from IyvDB (Optionmetrics). The database contains (annualized, continuously compounded) zero curves derived from LIBOR rates from the British Bankers Association (BBA) and settlement prices of Chicago Mercantile Exchange (CME) Eurodollar futures. For a given futures contract, the appropriate interest rate corresponds to the zero-coupon rate that has a maturity equal to the futures’ expiration date, and is obtained by linearly interpolating between the two closest zero-coupon rates on the zero curve.

Microstructure noise is a major concern when computing dividend prices from spot and futures data, a point well illustrated by Boguth et al. (2012). Since replication of dividend claims requires a contemporaneous long and short position in the underlying and futures contract, small pricing frictions in the prices of these primary instruments may propagate and generate measurement errors in synthetic dividend prices. These concerns are even more pressing for our analysis, which uses data from relatively illiquid contracts. In order to clean dividend prices of the noise component, we follow a three-step procedure. First, we construct a daily time series of dividend prices for each futures contract available in our sample. Second, we fit a Hodrick Prescott filter to each individual time series of dividend prices, using a smoothing parameter of $100 \times (252)^2$. Finally, we construct quarter-end prices of dividend claims with one year time to maturity by taking the cyclical component of the Hodrick-Prescott filtered dividend price of the fourth available contract (i.e., one year to

---

As of March 2012, available maturities include 11 contracts in the March cycle.
maturity). The prices we obtain are not tradable and cannot be viewed as real time trading signals (the filtering requires knowledge about the dividend prices implied by future spot and futures prices), but can be interpreted as the dividend price that would prevail after removing the effects induced by microstructure noise in the prices of the primary instruments.

Although our analysis focuses on dividend claims with one year to maturity, Figure 3.1 plots the time series of dividend prices with maturities of 0.25-1.00 years for completeness. As expected, dividend prices are increasing in maturity: this is a natural consequence of no arbitrage, since the payoffs of the dividend claims we construct consist of the cumulative dividends paid by the index up until the maturity of the corresponding futures contract. In order to gauge the effectiveness of our filtering technique, Figure 3.2 compares the time series of the 1-year dividend claim we construct with that based on options data computed by Binsbergen et al. (2012b): the two series feature a high degree of comovement.

### 3.3.2 Dividends

In order to avoid issues related to seasonality in index distributions, we construct annual dividends on the S&P500 by summing daily figures:

\[ D_t = \sum_{j=0}^{252} D_{t-j/252}, \]

so that annual (log) dividend growth is given by:

\[ \Delta d_{t+1} = \log \left( \frac{D_{t+1}}{D_t} \right). \]

S&P500 dividends can be computed by multiplying by the lagged price index the difference between the total return index and the price index (both available on Thomson Datastream). This aggregation procedure implicitly assumes that interim dividends are kept as cash rather than being re-invested in an interest bearing account or in the S&P500. The re-investment assumption has potential important implications for the empirical analysis, as discussed by Binsbergen & Kojien (2010). We choose not to use the market re-investment assumption in order to ensure that realized returns on dividend claims are not driven by risk premia of long term assets; since Binsbergen & Kojien (2010) show that there is no material difference
Figure 3.1: Dividend prices
This figure plots the time series of dividend prices (in S&P500 index points) for 0.25, 0.50, 0.75, and 1.00 years maturities.

between a strategy that does not re-invest dividends and one that invests dividends in risk free securities, we choose the former approach for simplicity.

3.3.3 Excess returns
The definition of index excess returns is standard:

\[ htr_{t+1,1} = \log \left( \frac{P_{t+1}}{P_t} \right) - y_t^{(1)}, \]

where \( P_{t+1} \) is either the S&P500 price or total return index (both from Thomson Datastream); the short term rate \( y_t^{(1)} \) is described below. Excess returns on the one year dividend claim,
Figure 3.2: Dividend prices: 1 year maturity
This figure plots two series of SP500 dividend prices (1 year maturity). Circles indicate prices computed following the methodology described in the data section, while crosses correspond to the prices constructed by Binsbergen et al. (2012b) (the data is available on the AEA website).

\[ strx_{t+1,1} = \log \left( \frac{D_{t+1}}{P_{t,1}} \right) - y_{t}^{(1)}. \]
3.3.4 Price to dividend ratios

We construct (log) price-to-dividend ratios on dividend claims and index as:

\[ \text{stpd}_{t,1} = \log \left( \frac{P_{t,1}}{D_t} \right) \]
\[ \text{ltpd}_{t,1} = \log \left( \frac{P_t}{D_t} \right). \]

The interpretation of price-to-dividend ratios of dividend claims is analogous to the classic one of Campbell (1991). When future dividends are known, dividend prices simplify to the value of dividends discounted by the risk free rate:

\[ P_{t,n} = \sum_{j=0}^{n} D_{t+j} e^{y_{t}^{(n)}}, \]

in general, however, dividends are uncertain and thus may command a risk premium. A simple accounting identity can be used to make the link between dividend prices and risk premia more precise:

\[ \text{stpd}_{t,n} = E_t \left[ \sum_{j=1}^{n} \Delta d_{t+j} \right] - E_t \left[ \sum_{j=1}^{n} \text{str} x_{t+j,n+1-j} \right] - E_t \left[ \sum_{j=1}^{n} y_{t+j}^{(1)} \right]. \]

This accounting identity is similar to the expression for equity yields derived by Binsbergen et al. (2012a): the prices of dividends are high relative to current dividends when expected dividend growth is high or expected returns are low.\(^4\) Given our focus on one-year horizons, the accounting identity simplifies to:

\[ \text{stpd}_{t,1} = E_t \left[ \Delta d_{t+1} \right] - E_t \left[ \text{str} x_{t+1,1} \right] - y_{t}^{(1)}. \] (3.9)

\(^4\)The identity can be derived as follows. The (log) holding period excess return on a dividend strip is given by

\[ \text{str} x_{t+1,n} = \log \left( \frac{P_{t+1,n-1}}{P_{t,n}} \right) - y_{t}^{(1)} \]
\[ = \log \left( \frac{STPD_{t+1,n-1}}{STPD_{t,n}} \right) \frac{D_{t+1}}{D_t} - y_{t}^{(1)} \]
\[ = \text{stpd}_{t+1,n-1} - \text{stpd}_{t,n} + \Delta d_{t+1} - y_{t}^{(1)}. \]

Re-arranging, one can obtain a recursive representation of normalized dividend prices as a function of discount
Figure 3.3 plots the time series of $stpd_t$ (since we focus on one-year maturities, we drop maturity subscripts from now on). The dynamics are consistent with intuition: price-to-dividend ratios plunge during the two financial crises in the sample (2001 and 2008), reflecting lower dividend growth expectations, higher discount rates, or a combination of both.

Table 3.1 reports the output from regressions of $\Delta d_{t+1}$ and $strx_{t+1}$ on $pd_t$. The slope coefficients of these two regressions can be interpreted as a decomposition of the variance of $stpd_t$, suggesting that more than half of the variation of the dividend prices can be attributed to time variation in expected returns.\(^5\) By way of comparison, Binsbergen et al. (2012a) use dividend swaps and find that equity yields predict dividend growth with an $R^2$ of 76% for 1 year horizons, and that over 70% of their variance is attributable to time varying dividend expectations. Our evidence, on the other hand, makes a stronger case for time-variation in the price of risk, rather than dividend growth predictability. This wedge between the results may be due to either the choice of sample (Binsbergen et al. (2012a) use data after October

rates and dividend growth rates

$$stpd_{t,n} = stpd_{t+1,n-1} + \Delta d_{t+1} - strx_{t+1,n} - y_t^{(1)},$$

subject to the terminal condition $STPD_{t,0} = 1$ ($stpd_{t,0} = 0$). Substituting forward:

$$stpd_{t,n} = \sum_{j=1}^{n} \Delta d_{t+j} - \sum_{j=1}^{n} strx_{t+j,n+1-j} - \sum_{j=1}^{n} y_t^{(1)}.$$

Taking expectations conditional on time-$t$ yields:

$$stpd_{t,n} = E_t \left[ \sum_{j=1}^{n} \Delta d_{t+j} \right] - E_t \left[ \sum_{j=1}^{n} strx_{t+j,n+1-j} \right] - E_t \left[ \sum_{j=1}^{n} y_t^{(1)} \right].$$

\(^5\)Multiplying both sides of 3.9 by $stpd_{t+1} - E[stpd_{t+1}]$ and taking expectations yields:

$$Var(stpd_{t+1}) = Cov(stpd_{t+1, \Delta d_{t+1}}) - Cov(stpd_{t+1, strx_{t+1}}) - Cov(stpd_{t+1, y_t^{(1)}}),$$

implying

$$1 = \frac{Cov(stpd_{t+1, \Delta d_{t+1}})}{Var(stpd_{t+1})} - \frac{Cov(stpd_{t+1, strx_{t+1}})}{Var(stpd_{t+1})} - \frac{Cov(stpd_{t+1, y_t^{(1)}})}{Var(stpd_{t+1})}. $$
2002, while our dataset covers a sample period starting in June 1992), or to differences in
the information content of dividend swap versus index futures.

Figure 3.3: Price-to-dividend ratio of dividend claim (1 year maturity)
This figure plots \( stpd_{t,1} = \log \left( \frac{P_{t+1}}{D_t} \right) \).

\[\begin{array}{c}
92 & 94 & 96 & 98 & 00 & 02 & 04 & 06 & 08 & 10 & 12 \\
-0.5 & -0.4 & -0.3 & -0.2 & -0.1 & 0 & 0.1 & 0.2 \\
\end{array}\]

date

3.3.5 Short term risk premia

We construct an estimate of short term risk premia by taking the fitted values from return
predictability regressions:

\[ strp_t = -0.01 - 0.57stpd_t. \]

Figure 3.4 plots the estimated STRP. The dynamics of STRP reveal sizable increases in risk
aversion during the busts of the Dotcom bubble and the collapse of Lehman brothers. The
Table 3.1: Information in stpd

This table reports the output from regressions $\Delta d_{t+1} = const. + \beta_{stpd_t} + \epsilon_{t+1}$ and $strx_{t+1} = const. + \beta_{stpd_t} + \epsilon_{t+1}$. T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted $R^2$ and the $\chi^2$ test statistic and p-value are also reported. Sample period: 1992 Q2 - 2012 Q1.

<table>
<thead>
<tr>
<th></th>
<th>const.</th>
<th>stpd</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta d_{t+1}$</td>
<td>0.03</td>
<td>0.43</td>
<td>43.29%</td>
<td>7.36</td>
</tr>
<tr>
<td></td>
<td>2.20</td>
<td>2.71</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>strx_{t+1}</td>
<td>-0.01</td>
<td>-0.57</td>
<td>53.70%</td>
<td>14.69</td>
</tr>
<tr>
<td></td>
<td>-0.41</td>
<td>-3.83</td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

The table also plots a survey-based estimate long term risk premia (LTRP) taken from Graham (2013). A comparison of the dynamics of STRP and LTRP reveals some interesting facts about the time variation in the price of risk across different horizons. First, STRP are on average higher than LTRP (the left scale is x20 the right scale), consistent with the findings of Binsbergen et al. (2012b). Second, STRP appear to be more persistent than LTRP. Third, the dynamics of the two series show a large degree of comovement: both STRP and LTRP peak in 2001 and 2009, and hit a trough in 2005. An exception to this overall pattern is the divergence that occurs from 2011 onwards: while STRP achieve new lows, LTRP hike up until levels witnessed in the aftermath of the collapse of Lehman brothers.

3.3.6 Bonds data

We construct bond yields and excess returns as follows. We obtain constant maturity Treasury yields (CMT) computed by the U.S. Treasury and published in the Federal Reserve Statistical Release H.15 report. We assume that yields can be treated as yields on par coupon bonds, and that only Treasuries with maturity greater than one year pay (semi-annual) coupons. We first obtain a set of equally spaced (semiannual frequency) yields to maturity by interpolating available yields with the Akima (1970) algorithm. Next we bootstrap to obtain a set of (simple, semiannual) zero coupon yields. Finally, we convert yields to
Figure 3.4: Risk premia
This figure plots the estimated STRP (in circles, left scale) and LTRP taken from Graham (2013) (crosses, right scale).

their continuously compounded form. The output of these data manipulations are time series of spot continuously compounded discount rates at evenly spaced maturities between 1 and 5 years \( y_t^{(n)} \), \( n = 1, \ldots, 5 \). Yields, which are related to bond (log) prices via \( p_t^{(n)} = -ny_t^{(n)} \), are then used to compute annual excess returns:

\[
br x_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)},
\]

and forward spreads:

\[
f_t^{(n)} - y_t^{(1)} = p_t^{(n-1)} - p_t^{(n)} - y_t^{(1)}.
\]
3.4 Empirical results

To investigate the link between STRP and LTRP empirically we run regressions of S&P500 realized excess returns on STRP:

$$\ln r_{t+h} = \text{const.} + \beta_h \text{strp}_t + \epsilon_{t+h},$$ (3.10)

where $h$ denotes the forecasting horizon; we consider quarterly horizons out to 1 year, $h = 0.25, 0.50, 0.75, 1.00$. When data sampling is finer than forecasting horizons, predictive regressions are overlapping. The overlap causes regression errors to be serially correlated even under the null hypothesis of no predictability: estimates of the covariance matrix of coefficients based on the classic assumption of homoskedasticity and no autocorrelation overstate the statistical significance of the estimates. In order to rule out spurious statistical significance, we estimate the covariance matrix of coefficients using Newey & West (1987).

Table 3.2 contains the output of the regressions for both cum-dividend (Panel A) and ex-dividend (Panel B) returns. The statistical significance of the loadings of index realized excess returns on STRP is high for all horizons: all t-statistics are above standard levels of statistical significance. Statistical significance tends to be decreasing in forecasting horizon. Adjusted $R^2$, on the other hand, are increasing in maturity: they range from 3.46% to 9.63%. Given the simplicity of the econometric model and the high noise in realized returns, the degree of predictability we find is interesting.

The most striking feature of the results relates to its economic significance. Since both regressands and regressors are log returns, the coefficients can be interpreted as elasticities. The magnitudes are economically significant. Importantly, they point in the direction of a negative correlation between STRP and LTRP: higher levels of STRP are associated with lower subsequent realizations of excess returns at the index level.
Table 3.2: Predictability of S&P500 returns: STRP
This table reports the output from regressions of S&P500 excess returns on a constant and STRP: \( \text{ltr}:x_{t+h} = \text{const.} + \beta_h \text{strp}_t + \epsilon_{t+1} \). Holding periods vary from 1 to 4 quarters: \( h = 0.25, 0.50, 0.75, 1.00 \). Excess returns are constructed from the S&P500 total return (Panel A) or price index (Panel B). T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted \( R^2 \) and the \( \chi^2 \) test statistic and p-value are also reported. Sample period: 1992 Q2 -2012 Q1.

<table>
<thead>
<tr>
<th></th>
<th>( h )</th>
<th>( \text{const.} )</th>
<th>( \text{strp} )</th>
<th>( \bar{R}^2 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td>0,25</td>
<td>0,01</td>
<td>-0,26</td>
<td>3,46%</td>
<td>7,65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,65</td>
<td>-2,77</td>
<td></td>
<td>0,01</td>
</tr>
<tr>
<td></td>
<td>0,50</td>
<td>0,03</td>
<td>-0,46</td>
<td>5,55%</td>
<td>4,43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,64</td>
<td>-2,11</td>
<td></td>
<td>0,04</td>
</tr>
<tr>
<td></td>
<td>0,75</td>
<td>0,04</td>
<td>-0,66</td>
<td>7,38%</td>
<td>3,77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,62</td>
<td>-1,94</td>
<td></td>
<td>0,05</td>
</tr>
<tr>
<td></td>
<td>1,00</td>
<td>0,05</td>
<td>-0,87</td>
<td>9,63%</td>
<td>4,59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,65</td>
<td>-2,14</td>
<td></td>
<td>0,03</td>
</tr>
<tr>
<td>Panel B</td>
<td>0,25</td>
<td>0,01</td>
<td>-0,25</td>
<td>3,21%</td>
<td>7,59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,04</td>
<td>-2,75</td>
<td></td>
<td>0,01</td>
</tr>
<tr>
<td></td>
<td>0,50</td>
<td>0,02</td>
<td>-0,45</td>
<td>5,16%</td>
<td>4,39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,04</td>
<td>-2,09</td>
<td></td>
<td>0,04</td>
</tr>
<tr>
<td></td>
<td>0,75</td>
<td>0,03</td>
<td>-0,64</td>
<td>6,88%</td>
<td>3,70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,02</td>
<td>-1,92</td>
<td></td>
<td>0,05</td>
</tr>
<tr>
<td></td>
<td>1,00</td>
<td>0,03</td>
<td>-0,84</td>
<td>9,03%</td>
<td>4,51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,04</td>
<td>-2,12</td>
<td></td>
<td>0,03</td>
</tr>
</tbody>
</table>
A natural question is whether STRP features explanatory power over and above that of market price to dividend ratios (LTPD), a standard predictor in the literature. These regressions have a strong theoretical motivation: under the assumption of rationality and no bubbles, it can be easily shown that LTPD vary over time only if either dividend growth or excess returns are predictable. Examining the joint predictive power of STRP and LTPD is interesting for two reasons. First, since the two variables capture time variation of future excess returns at different frequencies (we find that risk premia forecast excess returns over short horizons, while LTPD is known to have predictive power over long horizons), a model that includes both variables can lead to better forecasting performance. Second, by controlling for dividend yields we control for time-varying expectations about all future excess returns. Thus we can isolate better the marginal role of short term risk premia.

Table 3.3 summarizes the evidence about LTPD-predictability in our sample, by showing the output of forecasting regressions of realized excess returns on lagged LTPD. The results are consistent with the empirical literature. LTPD predicts excess returns with a negative sign (high prices relative to dividends capture low discount rates, rather than high dividend growth), and forecasting power builds over longer maturities: both the t-statistics and adjusted $R^2$ rise monotonically with the forecasting horizon.

Having run univariate regressions on STRP and LTPD separately, we analyze their combined predictive power in bivariate regressions:

$$ltrx_{t+h} = \text{const.} + \beta_{1h}\text{strp}_{t} + \beta_{2h}\text{ltpd}_{t} + \epsilon_{t+h}. \tag{3.11}$$

Table 3.4 summarizes the regression results. While the inclusion of LTPD seem to absorb some of the forecasting power, the statistical significance of STRP remains high: slope coefficients are significant at least at the 5% level across all horizons. Also the economic significance is only marginally affected: the slope coefficients remain negative. Importantly, STRP seems to absorb the forecasting power for horizons out to two quarters.

There are three main lessons to be learned from the results of regressions (3.10) and (3.11). First, the finding that higher STRP predict lower subsequent excess returns features high statistical and economic significance. The finding holds across different specifications of dividend expectations and is robust to the inclusion of LTPD. Second, while STRP and LTPD
Table 3.3: Predictability of S&P500 returns: LTPD

This table reports the output from regressions of S&P500 excess returns on a constant and LTPD: \( ltr_{t+h} = \text{const.} + \beta_h ltpd_t + \epsilon_{t+1} \). Holding periods vary from 1 to 4 quarters: \( h = 0.25, 0.50, 0.75, 1.00 \). Excess returns are constructed from the S&P500 total return (Panel A) or price index (Panel B). T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted \( R^2 \) and the \( \chi^2 \) test statistic and p-value are also reported. Sample period: 1992 Q2 - 2012 Q1.

<table>
<thead>
<tr>
<th></th>
<th>( h )</th>
<th>( \text{const.} )</th>
<th>( ltpd )</th>
<th>( \bar{R}^2 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td>0.25</td>
<td>0.25</td>
<td>-0.06</td>
<td>2.61%</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.75</td>
<td>-1.68</td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.53</td>
<td>-0.13</td>
<td>6.94%</td>
<td>4.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.15</td>
<td>-2.06</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.82</td>
<td>-0.20</td>
<td>10.98%</td>
<td>5.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.42</td>
<td>-2.30</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.11</td>
<td>-0.27</td>
<td>14.87%</td>
<td>6.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.64</td>
<td>-2.50</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Panel B</td>
<td>0.25</td>
<td>0.22</td>
<td>-0.05</td>
<td>2.05%</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.59</td>
<td>-1.55</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.49</td>
<td>-0.12</td>
<td>5.90%</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.97</td>
<td>-1.91</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.76</td>
<td>-0.19</td>
<td>9.53%</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.22</td>
<td>-2.14</td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.03</td>
<td>-0.25</td>
<td>13.08%</td>
<td>5.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.43</td>
<td>-2.34</td>
<td></td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table 3.4: Predictability of S&P500 returns: bivariate regressions
This table reports the output from regressions of S&P500 excess returns on a constant, STRP, and LTPD: \( ltr: r_{t+h} = const. + \beta_{1h} strp_t + \beta_{2h} ltpd_t + \epsilon_{t+1} \). Holding periods vary from 1 to 4 quarters: \( h = 0.25, 0.50, 0.75, 1.00 \). Excess returns are constructed from the S&P500 total return (Panel A) or price index (Panel B). T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted \( R^2 \) and the \( \chi^2 \) test statistic and p-value are also reported. Sample period: 1992 Q2 - 2012 Q1.

<table>
<thead>
<tr>
<th></th>
<th>( h )</th>
<th>const.</th>
<th>strp</th>
<th>ltpd</th>
<th>( R^2 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td>0.25</td>
<td>0.19</td>
<td>-0.21</td>
<td>-0.04</td>
<td>4.16%</td>
<td>19.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.42</td>
<td>-2.22</td>
<td>-1.33</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.43</td>
<td>-0.34</td>
<td>-0.10</td>
<td>9.17%</td>
<td>18.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.81</td>
<td>-2.01</td>
<td>-1.71</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.69</td>
<td>-0.47</td>
<td>-0.16</td>
<td>13.84%</td>
<td>18.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.11</td>
<td>-1.93</td>
<td>-1.98</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.93</td>
<td>-0.61</td>
<td>-0.22</td>
<td>18.70%</td>
<td>24.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.44</td>
<td>-2.28</td>
<td>-2.29</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Panel B</td>
<td>0.25</td>
<td>0.16</td>
<td>-0.21</td>
<td>-0.04</td>
<td>3.54%</td>
<td>18.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.24</td>
<td>-2.19</td>
<td>-1.19</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.39</td>
<td>-0.34</td>
<td>-0.09</td>
<td>8.04%</td>
<td>16.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.62</td>
<td>-1.97</td>
<td>-1.56</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.62</td>
<td>-0.46</td>
<td>-0.15</td>
<td>12.28%</td>
<td>16.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.90</td>
<td>-1.88</td>
<td>-1.82</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.85</td>
<td>-0.60</td>
<td>-0.21</td>
<td>16.78%</td>
<td>21.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.20</td>
<td>-2.21</td>
<td>-2.10</td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

58
feature some commonality, marginal significance statistics suggest that they carry different information. Third, differences in the patterns of loadings across maturities suggest that time variation in STRP and LTPD captures high and low, respectively, frequency components of future realized returns.

To the extent that short term risk premia capture investors’ risk aversion for low duration investments, it is reasonable to conjecture that they possess forecasting ability for securities that have lower duration than equities. We test this conjecture by looking at the ability of STRP to predict excess returns on Treasury securities. We consider Treasury securities for two reasons. First, bonds have a fixed maturity, so the decomposition of their returns into short- and long-term shocks to discount factors is easier than for equities. Second, the absence of cash flow risk allows us to focus on the “preference for duration” component of short term risk premia.

\[ brx_{t+h} = \text{const.} + \beta_{h} \text{str}p_{t} + \varepsilon_{t+h}. \]  

(3.12)

Since \( n \)-year forward spreads are known to have predictive power for the annual returns of \( n \)-year bonds (Fama & Bliss (1987)), we also estimate a specification that includes forward spreads as controls:

\[ brx_{t+h} = \text{const.} + \beta_{1h} \text{str}p_{t} + \beta_{2h} \left( f^{(h)}_{t} - y^{(1)}_{t} \right) + \varepsilon_{t+h}. \]  

(3.13)

Tables 3.5 and 3.6 contain the results of the predictive regressions. For 2-years bonds, the loadings on STRP feature high statistical significance, while the null hypothesis of no predictability cannot be rejected for longer maturities. This effect is robust to controlling for forward spreads. Importantly, contrary to what we find in the equity forecasting regressions, the loading on STRP is positive. This finding is consistent with the traditional intuition that high ex-ante risk premia should command, on average, high realized excess returns; at the same time, however, it is suggestive that the behavior of short duration assets may be distinct from that of long duration assets.

3.5 Understanding predictability

The finding that STRP predict realized excess returns with a negative sign is puzzling as it suggests that higher levels of the price of short term risk tend to be accompanied by lower
Table 3.5: Bond predictability: univariate regressions

This table reports estimates from OLS univariate regressions of n-year Treasury securities annual excess returns on STRP: \( brx_{t+1}^{(n)} = const. + \beta_h strp_t + \epsilon_{t+h} \). The bond maturity \( n \) ranges from two to five years. T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted \( R^2 \) and the \( \chi^2 \) test statistic and p-value are also reported. Sample period: 1992 Q2 - 2012 Q1.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( const. )</th>
<th>( strp )</th>
<th>( \bar{R}^2 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.06</td>
<td>8.71%</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>3.46</td>
<td>2.23</td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.07</td>
<td>2.64%</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td>1.58</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.13%</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>3.12</td>
<td>0.89</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>0.03</td>
<td>0.02</td>
<td>-1.18%</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>3.33</td>
<td>0.27</td>
<td></td>
<td>0.79</td>
</tr>
</tbody>
</table>
Table 3.6: Bond predictability: bivariate regressions

This table reports estimates from OLS univariate regressions of n-year Treasury securities annual excess returns on STRP and forward spreads: $brx_{t+4}^{(n)} = const. + \beta_{1h}strp_t + \beta_{2h}(f_t^{(n)} - y_t^{(1)})\epsilon_{t+h}$. The bond maturity $n$ ranges from two to five years. T-statistics, reported below the point estimates, use Newey & West (1987) standard errors (6 quarters lag). The adjusted $R^2$ and the $\chi^2$ test statistic and p-value are also reported. Sample period: 1992 Q2 - 2012 Q1.

<table>
<thead>
<tr>
<th>$n$</th>
<th>const.</th>
<th>strp</th>
<th>$f_t^{(n)} - y_t^{(1)}$</th>
<th>$\bar{R}^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0,01</td>
<td>0,06</td>
<td>0,19</td>
<td>8,28%</td>
<td>4,76</td>
</tr>
<tr>
<td></td>
<td>1,64</td>
<td>2,15</td>
<td>0,43</td>
<td></td>
<td>0,09</td>
</tr>
<tr>
<td>3</td>
<td>0,01</td>
<td>0,07</td>
<td>0,34</td>
<td>2,76%</td>
<td>2,32</td>
</tr>
<tr>
<td></td>
<td>1,33</td>
<td>1,52</td>
<td>0,61</td>
<td></td>
<td>0,31</td>
</tr>
<tr>
<td>4</td>
<td>0,01</td>
<td>0,06</td>
<td>0,51</td>
<td>1,28%</td>
<td>1,31</td>
</tr>
<tr>
<td></td>
<td>1,27</td>
<td>0,91</td>
<td>0,93</td>
<td></td>
<td>0,52</td>
</tr>
<tr>
<td>5</td>
<td>0,02</td>
<td>0,03</td>
<td>0,67</td>
<td>2,12%</td>
<td>1,65</td>
</tr>
<tr>
<td></td>
<td>1,23</td>
<td>0,38</td>
<td>1,28</td>
<td></td>
<td>0,44</td>
</tr>
</tbody>
</table>
realized returns of long duration assets. In this section we investigate whether these results can be reconciled with any of the leading asset pricing models. We evaluate the implications of three asset pricing models: the reduced form model of Lettau & Wachter (2007), consumption habit (Campbell & Cochrane (1999)), and long run risk (Bansal & Yaron (2004)). The main conclusion of the section is that none of these models is able to generate the predictability patterns that we find in the data.

We start by considering the implications of the model by Lettau & Wachter (2007), introduced above as a motivating example. What range of regression coefficients does the model imply? Equation (3.6) provides a means to answer this question. Unfortunately, the expression requires taking the derivative of a complex function involving non-linearities and an infinite series, so an analytical expression for β cannot be found. Even though we cannot pin down the magnitude of β, we can say something about the loading of expected returns of individual dividend strips on ||σ_d||x_t via (3.3). Computation of (3.3) requires a parametrization of the model. We follow Lettau & Wachter (2007) in part, by considering annual frequency and setting r_f = 1.93%, g = 2.28%, φ_x = 0.91, φ_z = 0.87, x̄ = 0.625, ||σ_d|| = 0.145, ||σ_z|| = 0.0032, ||σ_x|| = 0.24. We depart from Lettau & Wachter (2007) in the parametrization of σ_d, σ_z, and σ_x. Lettau & Wachter (2007) assume that dividend shocks are correlated with shocks to dividend expectations, but not with shocks to the price of risk. These assumptions, together with the estimates of the conditional volatilities, imply the following specification for Σ = [σ_d' σ_z' σ_x']’

\[
\begin{pmatrix}
σ_{dd} & 0 & 0 \\
σ_{zd} & σ_{zz} & 0 \\
0 & 0 & σ_{xx}
\end{pmatrix},
\]

where

\[
σ_{dd} = ||σ_d||
\]
\[
σ_{zd} = ρ_{Z,d}||σ_z||
\]
\[
σ_{zz} = \sqrt{||σ_z||^2 - ρ_{Z,d}^2||σ_z||^2}
\]
\[
σ_{xx} = ||σ_x||.
\]
\[ \rho_{x,d} = \frac{\sigma_x' \sigma_d}{\|\sigma_x\| \|\sigma_d\|}. \]

As Lettau & Wachter (2007) note, the assumption of no conditional correlation between dividend and price of risk shocks is not innocuous and implies assumptions about more primitive features of the economy. Instead of choosing a particular specification, we explore model predictions for a variety of combinations of \( \rho_{zd} \) and \( \rho_{xd} \). The characterization differs from Lettau & Wachter (2007) only in the choice of \( \sigma_x \)

\[ \sigma_x = [\sigma_{zd} \quad 0 \quad \sigma_{xx}], \]

where

\[ \sigma_{zd} = \rho_{x,d} \|\sigma_x\| \]
\[ \sigma_{xx} = \sqrt{\|\sigma_x\|^2 - \rho_{x,d}^2 \|\sigma_x\|^2}. \]

Not all combinations of \( \rho_{zd} \) and \( \rho_{xd} \) are allowed: restrictions have to be imposed to ensure that \( P_t \) converges for all \( z_t \) and \( x_t \) and that \( \sigma_{xz} \) and \( \sigma_{xx} \) are real. Lettau & Wachter (2007) show that the price of equity converges to a finite value if and only if the following three conditions are satisfied:

1. \( |\phi_z| < 1 \).
2. \( |\phi_z - \sigma_x' \sigma_{zd} / \|\sigma_d\| | < 1 \).
3. \( -r + g + \tilde{B}_z (1 - \phi_z) \bar{x} + \frac{1}{2} \tilde{V} \tilde{V}' < 0 \).

where

\[ \tilde{B}_z = -\left( \frac{\sigma_d + \frac{\sigma_x' \sigma_{zd}}{\|\sigma_d\|}}{1 - \left( \phi_z - \frac{\sigma_x \sigma_{zd}}{\|\sigma_d\|} \right)} \right), \]

and

\[ \tilde{V} = \sigma_d + \frac{\sigma_x}{1 - \phi_z} + \tilde{B}_z \sigma_x. \]

Condition 1 is satisfied because we set \( \phi_z = 0.91 \). Condition 2 can be rewritten as follows

\[ |\phi_z - \rho_{zd} \|\sigma_x\|| \|\sigma_x\| < 1 \rightarrow \left| \rho_{zd} - \frac{\phi_z}{\|\sigma_x\|} \right| < \frac{1}{\|\sigma_x\|}. \]
which implies the following condition for $\rho_{zd}$

$$\rho_{zd} = \left( \frac{\phi_x}{||\sigma_z||} - \frac{1}{||\sigma_z||}, \frac{\phi_x}{||\sigma_z||} + \frac{1}{||\sigma_z||} \right).$$

For the set of admissible $\rho_{zd}$ derived above, we evaluate condition 3 numerically and rule out the $\rho_{zd}$ that do not satisfy it. We assume that $\Sigma = [\sigma'_d \quad \sigma'_z \quad \sigma'_{zd}]'$ equals:

$$\begin{pmatrix} \sigma_{dd} & 0 & 0 \\ \sigma_{zd} & \sigma_{zz} & 0 \\ \sigma_{zd} & 0 & \sigma_{xx} \end{pmatrix}.$$  

The unknown parameters can be computed from the knowledge of the conditional volatilities and pairwise correlations. We have:

1. $\sigma_{dd} = ||\sigma_d||$.
2. Since $\rho_{zd} = \frac{\sigma_z \sigma_{zd}'}{||\sigma_z|| ||\sigma_d||} = \frac{\sigma_{zd}}{||\sigma_z||}$, we have $\sigma_{zd} = \rho_{zd} ||\sigma_z||$.
3. $\sigma_{zz} = \sqrt{||\sigma_z||^2 - \sigma_{zd}^2}$.
4. The same argument as point 2 leads to: $\sigma_{zd} = \rho_{zd} ||\sigma_z||$.
5. $\sigma_{xx} = \sqrt{||\sigma_x||^2 - \sigma_{zd}^2}$.

Points 3 and 5 show that $\rho_{zd}$ and $\rho_{zd}$ must be such that the argument of the square roots is positive: we rule out all combinations of correlations that do not satisfy this condition.

Figure 3.5 shows which combinations of correlation coefficients ensure the convergence for our choice of parameters. It is interesting to note that for this specific choice of parameter values, the binding constraint comes in the form of a lower bound for $\rho_{zd}$ around $-0.3$. In figure 3.6, each panel shows the loadings on $x_t$ of expected returns (equation (3.3)) for a fixed maturity and different combinations of $\rho_{zd}$ and $\rho_{zd}$. The figures exhibit clear patterns. The key source of variation in the loadings is the correlation between shocks to the price of risk and to realized dividends. For a fixed combination of $\rho_{zd}$ and $\rho_{zd}$, the loadings barely change across maturities. Fixing maturity, the loadings decrease from approximately 0.8 for high values of $\rho_{zd}$, to approximately 0.2 for low values of $\rho_{zd}$. The sign of the loadings is
positive for all maturities: an increase in the price of risk raises expected excess returns at all horizons. Since the expected return on the index is a weighted sum of the expected returns on its dividend strips, the graphs suggest that a regression of realized excess returns on risk premia should yield a positive slope coefficient.

**Figure 3.5: Admissible correlations**
The region on the right is the set of \((\rho_{x,d}, \rho_{z,d})\) that ensure that: i) the price to dividend ratio of dividend strips converges for any maturity, and ii) the values of \(\sigma_{zz}\) and \(\sigma_{xz}\) are real.

![Admissible correlations](image)

The model explored by Bansal & Yaron (2004) implies that STRP are zero: long run risk models rule out the possibility that STRP has any predictive power with regards to future returns. This can be shown with straightforward algebra. The dynamics of long run risk economies can be described by the processes for (log) dividend growth, (log) consumption
Figure 3.6: Loadings of expected returns on $x_t$
These graphs illustrate the loadings of expected returns of dividend strips of maturity $N$ on the price of risk $x_t$ for a variety of $(\rho_{x,d}, \rho_{x,d})$ combinations. The lines are contours connecting the $(\rho_{x,d}, \rho_{x,d})$ pairs with the same loading $x_t$; the number on each contour indicates its level.

growth, the conditional expected growth rate, and economic uncertainty:

$$\Delta d_{t+1} = \mu_d + \phi z_t + w_d \sigma_d \epsilon_{d,t+1}$$

$$\Delta c_{t+1} = \mu_c + z_t + w_c \sigma_c \epsilon_{c,t+1}$$
\[ z_{t+1} = \phi z_t + w_x \sigma_t \epsilon_{x,t+1} \]
\[ \sigma^2_{t+1} = \mu_\sigma + \nu (\sigma^2_t - \sigma^2) + w_\sigma + \epsilon_{\sigma,t+1}, \]

where \( \epsilon_{d,t+1}, \epsilon_{c,t+1}, \epsilon_{z,t+1}, \) and \( \epsilon_{\sigma,t+1} \) are i.i.d. \( N(0,1) \). The equilibrium log return on the market asset is given by

\[ r_{t+1} = \Delta d_{t+1} + k_1 A_{1,m} z_{t+1} - A_{1,m} z_t + k_{1,m} A_{2,m} \sigma_{t+1}^2 - A_{2,m} \sigma_t^2, \]  

(3.14)

where a complete characterization of \( A_{1,m}, A_{2,m}, k_1 \) and \( k_{1,m} \) is provided by Bansal \& Yaron (2004). The risk premium on the first dividend claim is given by \( \log E_t[D_{t+1}] - \log P_{t,1} - r^f_t \). Let \( M_{t+1} \) denote the stochastic discount factor between \( t \) and \( t+1 \), i.e. \( P_{t,1} = E_t[M_{t+1}D_{t+1}] \). By the properties of normal variables, we have

\[ \log E_t[D_{t+1}] = \log D_t E_t[\exp(\Delta d_{t+1})] \]
\[ = d_t + E_t[\Delta d_{t+1}] + 0.5 Var_t[\Delta d_{t+1}] \]

\[ \log P_{t,1} = \log D_t E_t[\exp(m_{t+1} + \Delta d_{t+1})] \]
\[ = d_t + E_t[m_{t+1}] + E_t[\Delta d_{t+1}] \]
\[ + 0.5(Var_t[m_{t+1}] + Var_t[\Delta d_{t+1}] + 2Cov_t[m_{t+1}, \Delta d_{t+1}]) \]

\[ r^f_t = - \log E_t[\exp(m_{t+1})] \]
\[ = - E_t[m_{t+1}] - 0.5 Var_t[m_{t+1}]. \]

Hence, the expression for STRP simplifies to

\[ \log E_t[D_{t+1}] - \log P_{t,1} - r^f_t = -2Cov_t[m_{t+1}, \Delta d_{t+1}], \]

i.e. the risk premium depends on the covariance between \( m_{t+1} \) and \( \Delta d_{t+1} \). The innovations to the SDF are equal to

\[ m_{t+1} - E_t[m_{t+1}] = \lambda_m \sigma_t \epsilon_{c,t+1} - \lambda_{mz} \sigma_t \epsilon_{x,t+1} - \lambda_{ms} \sigma_t \epsilon_{\sigma,t+1}, \]
where expressions for the prices of risk $\lambda_{mc1}, \lambda_{mc2}, \lambda_{ma}$ are given in the original chapter. Since innovations to $\Delta d_{t+1}$ are equal to

$$\Delta d_{t+1} - E_t[\Delta d_{t+1}] = w_d \sigma_d \epsilon_{d,t+1},$$

it follows that $Cov_t[m_{t+1}, \Delta d_{t+1}] = 0$ and $sterp_t = 0$.

The intuition of this result is simple. In long run risk models, agents are fearful about states of the world in which expected growth prospects are low and economic uncertainty is high. A short maturity dividend claim is exposed to the risk that realized dividends undershoot expectations, but it is virtually not affected by long run growth prospects and economic uncertainty. Since shocks to realized dividends are not priced in this economy, agents demand zero risk premia to hold short maturity dividend claims.

Also a simple habit formation model will find it difficult to explain the data. As habit implies that dividend shocks are negatively correlated with innovations to the price of risk, the model can be (somewhat) casted in the context of Lettau & Wachter (2007) specification with $\rho_{ad} < 0$. In this case, however, the lower the correlation between realized dividends and the price of risk, the higher the implied slope coefficient.

The asset pricing models presented above fail to capture the predictability we find in the data. The mechanics of the failures are precious because they shed light on the direction of improvement that future models will have to pursue to match the empirical evidence of short term predictability. Bansal & Yaron (2004) on one side, and Lettau & Wachter (2007) and Campbell & Cochrane (1999) on the other, fail to capture predictability for two different reasons. Long run risk models deny the existence of predictability conditional on STRP because short term dividends are not exposed to the two sources of priced risk of the economy, i.e. long run economic uncertainty and growth prospects. The reduced form model by Lettau & Wachter (2007), on the other hand, does allow for predictability conditional on STRP, but implies a positive slope coefficient. The intuition behind this result is simple. Expected returns are increasing in the price of risk, which is revealed, up to a scaling factor, by the expected returns on the short term asset (via equation (3.3)). If the price of risk is the only state variable driving the term structure of expected returns and follows a persistent process, an increase in short term risk premia (i.e. the price of risk) will raise expected

68
excess returns at all horizons, yielding a positive slope coefficient in regressions of realized returns on STRP. Hence, we learn that two ingredients have to co-exist in a model for it have a chance to rationalize our empirical findings. First, shocks to realized dividends must be priced. Second, the term structure of risk premia must be driven by a multivariate state vector. Moreover, these additional priced state variables affect the short and long-end of the term structure of risk premia in ways that are weakly (if not negatively) correlated.

The result is potentially consistent with vast literature that studies the potential role of economic uncertainty on asset pricing.\(^6\) Short-term dividends have an advantage with respect to longer-term cash-flow: they are less uncertain. Indeed, they are often announced in advance and more substitutable to short-term bonds. In periods of heightened uncertainty, we see the emergence of a preference for short-term duration assets and it is possible that uncertainty-averse agents prefer short duration assets (short-term dividend claims) to the expense of long-term duration assets (equity shares). Depending on the extent to which agents demand a reduction in the cash-flow duration of their exposure, this may give rise to a negative slope coefficient in our long-term predictive regressions.

### 3.6 Conclusion

We construct short horizon risk premia by combining dividend prices with dividends expectations, and find evidence that risk premia capture short horizon predictability in S&P500 realized excess returns. The results are robust to the inclusion of dividend yields in the conditioning information set. The slope coefficients of the predictive regressions are difficult to reconcile with the predictions of leading asset pricing models. The failure of these models may be the consequence of their inability to produce a multi-factor structure for the term structure of equity risk premia.

---

\(^6\)See Miller (1977) and Diether et al. (2002) for behavioral explanations, and Buraschi et al. (2013) for a general equilibrium study of economies with economic uncertainty.
Chapter 4

Monetary policy and bond risk premia\textsuperscript{1}

4.1 Introduction

This chapter investigates the extent to which investors view monetary policy shocks as a source of priced risk in bond markets, and whether this channel is empirically important to explain the time-variation in risk premia. Monetary policy and bond prices are connected via two channels. First, central banks use the nominal short-term rate as the primary instrument of monetary policy. Second, absence of arbitrage opportunities implies that bond yields reflect risk neutral expectations about future short rates. Hence, both institutional features (short rate as a policy instrument) and economic restrictions (no arbitrage) enforce a fundamental link between monetary policy and the entire term structure of interest rates. In fact, a large literature documents strong responses in yields to news about monetary policy (see, for instance, Kuttner (2001), Piazzesi & Fleming (2005), and Gurkaynak et al. (2005)). Little is known, however, about whether policy influences current yields through physical expectations about future short rates or, rather, risk premia. In this chapter, we use a new approach to identify monetary policy path shocks and investigate the link between monetary policy and bond risk premia.

\textsuperscript{1}The work in this chapter is based on and extends the December 2012 version of Buraschi et al. (2012). I am indebted to Marcin Kacperczyk and Alessandro Beber for many useful comments and suggestions.
From an empirical point of view, there is little agreement in the literature on how to measure monetary policy actions. A first difficulty is related to the fact that a significant component of policy actions reflects the systematic response of the policy instrument to the macro-economic environment, rather than authentically exogenous policy shocks. In practice, researchers must make identifying assumptions to be able to disentangle the systematic component from the monetary policy shock. Needless to say, the dynamic properties of the resulting series are highly dependent on these assumptions (Christiano et al. (1999)). A second difficulty is represented by the fact that data on short-term target changes is unlikely to capture the richness of policy decisions. For instance, market participants may fully foresee target rate changes, but be considerably surprised about the path of future policy as inferred from the statements of the members of the policy committee: in these circumstances, a measure of monetary policy shocks based on the policy instrument may significantly underestimate the extent of exogenous variation in monetary policy. This concern is particularly important for our setting, since monetary policy is known to influence bond yields more via path, rather than target rate, surprises (Gurkaynak et al. (2005)); this possibility is also widely understood by policymakers. In this article, we propose a novel measure of monetary policy shocks that attempts to address both challenges.

The theoretical motivation of why monetary policy may affect the dynamics of bond risk premia can be understood both in the context of economies with a time-varying price of risk and in economies with a time-varying quantity of risk. In habit economies, for instance, monetary policy may be important due to the effect of money non-neutrality on the dynamics of consumption surplus. In these economies, the marginal utility of agents depends on the distance of consumption from habit: agents become relatively more risk averse when

---

2See, for instance, Bernanke (2004b): 'The current funds rate imperfectly measures policy stimulus because the most important economic decisions, such as a family's decision to buy a new home or a firm's decision to acquire new capital goods, depend much more on longer-term interest rates, such as mortgage rates and corporate bond rates, than on the federal funds rate. Long-term rates, in turn, depend primarily on the current funds rate but on how financial market participants expect the funds rate and other short-term rates to evolve over time.'

3Money non-neutrality indicates the idea that monetary policy influences not only nominal quantities such as inflation and exchange rates, but also real aggregates such as employment, GDP, and consumption (see King (2000) and Christiano et al. (1999), for a review).
consumption drops relative to its recent trend. Hence, if monetary policy affects consumption (if money is non-neutral), it also affects marginal utility and the level of risk aversion of investors. To the extent that contractionary policies have a negative impact on consumption, a (positive) monetary policy shock will increase the price of risk as consumption falls closer to the habit stock.\footnote{The theoretical literature has studied this channel either in the context of internal habit formation, as in Sundaresan (1989), Constantinides (1990), and Heaton (1995), or in external habit specifications, as in Campbell & Cochrane (1999), Wachter (2006), and Buraschi & Jiltsov (2007)}. The link between monetary policy and risk premia can be also understood in the context of long run risk economies (see, for instance, Bansal & Yaron (2004) and Bansal & Shaliastovich (2013)). Since predictability is an implication of heteroskedastic fundamentals in long run risk economies, monetary policy can generate time varying risk premia in this class of models if it has an impact on the volatility of GDP growth and inflation. Given the large debate about the significance of monetary policy as a major determinant of economic (in)stability (see, for instance, Bernanke (2004a)), the long run risk perspective is, next to habit models, an insightful channel to understand the potential link between monetary policy and risk premia. This chapter, after quantifying the extent to which monetary policy is a source of variation for bond risk premia, empirically investigates the relative importance of the quantity versus price of risk channels as alternative potential sources of this time variation.

The results of the chapter can be summarised as follows. First, we explore a new strategy to identify forward-looking monetary policy shocks. In particular, we construct an empirical proxy of monetary policy path shocks from the residuals of a Taylor rule estimated on survey expectations. Central to this analysis is a new data set that includes joint expectations about the target rate (fed fund rate) and economic fundamentals (gdp and inflation). This data set is compiled at monthly frequency and is available in a panel data format so that, for each individual, we have observations on the expected counterparts to both the left and right-hand side of the Taylor rule. Importantly, this dataset allows us to empirically identify a measure of monetary policy path shocks, without having to make assumptions about the data generating process in the mind of the agents. While truly forward-looking and robust to model misspecification and estimation error, survey forecasts may however be plagued by other types of reporting biases. In order to ensure that the reporting biases
are quantitatively negligible, we conduct a host of quality checks. When we compare the properties of subjective macro expectations to those obtained from traditional econometric benchmarks, we find that the errors of consensus forecasts are in absolute value lower than their econometric counterparts. This is especially true for the Fed fund rates forecasts. This result is interesting and highlights the potential importance of structural breaks in the conduct of policy decisions. Another advantage of the panel structure of the data is that it circumvents the need to use pre-aggregated consensus data, which may in itself bias the results. Indeed, we compare models based on pre-aggregated consensus data, or panel data, using (i) pooled OLS, (ii) fixed effects, (iii) random effects, and find that a panel data approach with fixed effects is preferable over procedures based on pre-aggregated consensus data.

Second, we find that monetary policy path shocks represent an empirically important source of time variation in bond risk premia. We run predictability regressions of bond excess returns onto lagged monetary policy shocks and find that a one standard deviation increase in monetary policy shocks predicts an increase in future excess returns by, approximately, 0.4 standard deviations. Also the statistical significance is high: the null hypothesis that the slope coefficient is zero cannot be rejected at standard levels of significance, and monetary policy shocks alone account for approximately 15% of the variance of bond excess returns. We show that the evidence: (i) is robust across a variety of Taylor rule specifications; (ii) is present both in the full sample and in a subsample that excludes the last financial crisis; and (iii) survives after controlling for levels of macroeconomic activity.

Third, our proxy of monetary policy path shocks features strong co-movement with the risk premium factors based on yield curve information proposed by Cochrane & Piazzesi (2005) and Le & Singleton (2013). Univariate regressions of monetary policy path shocks onto the factor of Cochrane & Piazzesi (2005) or Le & Singleton (2013) yield $R^2$ of up to 21% and 32%, respectively, depending on the specification. In order to confirm the robustness of this interpretation, we construct a factor mimicking portfolio and quantify the extent to which monetary policy path shocks help explain the cross section of equity returns, after controlling for the Fama and French factors. Indeed, we find that the price of monetary policy path risk estimated in the second stage regressions ranges from 3.63% to 5.39% on an annualized basis, an economic magnitude that is second only to the market risk premium.
This result highlights that monetary policy path shocks have pervasive effects in both bond and equity markets.

Last, we investigate whether the link between monetary policy path shocks and bond risk premia can be understood in the context of time varying price, or quantity, of risk. In order to study the first channel channel, we assess the forecasting ability of path shocks for a proxy of consumption surplus. Next, we construct estimates of the conditional volatility of inflation and real economic growth, and ask whether path shocks are contemporaneously correlated with macroeconomic uncertainty. While the link with consumption surplus is weak, we find that regressions of real economic uncertainty onto path shocks yield $R^2$ around $R^2 = 20.5\%$. These results are consistent with a long run risk interpretation (quantity of risk channel), and suggest that monetary policy shocks make their way into risk premia by affecting the overall level of macroeconomic uncertainty (see Bansal, Kiku, and Amir Yaron (2012)).

The rest of the chapter is organized as follows. After a survey of related literature in the next subsection, Section 4.2 describes the data. Section 4.3 reviews the details behind the construction of monetary policy shocks, while Section 4.4 presents the core evidence on bond return predictability. Next, section 4.5 investigates whether the evidence is consistent with a price of risk, or quantity of risk channel. Section 4.6 concludes.

4.1.1 Related literature

This chapter is related to two streams of the literature. A first stream studies the implications of monetary policy announcements on realized returns and volatility of asset prices. Some of these empirical studies use the FOMC Federal Open Market Committee) meetings to identify releases of information about the conduct of monetary policy. The use of announcements is interesting because it allows to identify shocks to the information set of agents and understand the transmission mechanism of monetary policy. A classic study is Kuttner (2001), who disentangles the expected and unexpected components of monetary policy from federal funds futures and finds that the latter accounts for a large fraction of the daily variation of bond yields around FOMC announcements. He also finds, however, that the fraction of variance that can be accounted for by monetary policy surprises declines in maturity. Thus, it is not clear to what extent the fluctuations observed during the release of FOMC announcement
can be entirely attributed to policy news. Piazzesi & Fleming (2005) and Gurkaynak et al. (2005) tackle this puzzle from two different perspectives. Piazzesi & Fleming (2005) find that yield changes depend not only on the surprises, but also on the slope of the term structure. A possible interpretation of this result is that a positive (flat) slope may indicate fears about inflation (hard landing), so that a surprise increase (decrease) in the federal fund rate may translate into lower (higher) expected inflation, thus sending long rates in the opposite direction of the short rate. Gurkaynak et al. (2005), on the other hand, start from the intuition that surprises about the target rate represent only part of the news on announcement days: market participants are likely to update their beliefs about the future path of policy as well. Gurkaynak et al. (2005) construct a factor that proxies for news about the future path of policy, and find that it can explain much of the variance of long term yields that is unaccounted for by surprises about the federal funds target. More recently, Andersson (2007) extends the analysis to the euro area and documents that while bond yields feature an upsurge in volatility around announcements in both US and the euro-area, the effect is more pronounced in US following Federal Reserve monetary policy decisions. Overall, the previous studies indicate that monetary policy affects yields in a quantitatively significant fashion. These results, however, are not limited to bond markets and first moments. Bernanke & Kuttner (2005) find that an unanticipated cut in the Fed rate of 25 basis points triggers a 1% increase in the stock market; Lucca & Moench (2011) document that as much as 80% of the US equity premium is earned in the 24 hours before FOMC announcements. Beber & Brandt (2006) analyze the effect of announcements on the state price densities of bond prices implied by Treasury options. They find that all announcements decrease the second moment of the state price density, while the effect on higher moments depends on the sign of the news. 5 Importantly, both Bernanke & Kuttner (2005) and Beber & Brandt (2006) decompose the effect of announcements into their economic drivers and assess the quantitative relevance of the component due to changes in expected returns. Bernanke & Kuttner (2005) show that monetary policy shocks account for a high fraction of discount rate news; Beber & Brandt (2006), on the other hand, conclude that the results are strongly suggestive of counter-cyclical variation in relative risk aversion.

5Since the authors not only consider FOMC announcements, but also 9 additional US macroeconomic news releases, it is not possible to know how results would be affected if only FOMC announcements were taken into account.
A second stream of the literature studies the link between monetary policy and bond pricing in the context of no-arbitrage term structure models. There are two popular modeling strategies. A first strategy reconciles reduced-form short rate-term structure models with the notion that most modern monetary authorities implement policy by controlling the path of the short rate. Under no arbitrage, yields are risk neutral expectations of future short rates, so that, from the perspective of no-arbitrage term structure models, introducing monetary policy amounts to specifying the mapping between short rates and state variables in a fashion that is consistent with the policy response function of the central bank. Some notable applications of this strategy are Piazzesi (2005), Ang et al. (2007), and Chun (2011). Piazzesi (2005) obtains pricing implications for a flexible characterization of the jump process followed by the target rate; Ang et al. (2007) explore the restrictions implied by a host of Taylor rule specifications; Chun (2011) studies the link between inflation and GDP forecasts and bond yields by incorporating survey data into a term structure model via a forward looking Taylor rule. A second strategy introduces monetary policy by modeling inflation endogenously. The nominal pricing kernel is given by the real pricing kernel, deflated by the inflation rate. For a given choice of the real pricing kernel, the assumption that the central bank controls the nominal short rate implies a restriction on the (endogenous) inflation rate: the inflation process can be solved for by requiring consistency between the policy rule of the central bank and the Euler equation for the nominal short rate. This strategy is employed by Gallmeyer et al. (2007a) and Gallmeyer et al. (2007b), who obtain bond pricing implications by considering a simple contemporaneous Taylor rule specification in the context of a long run risk and habit formation economy, respectively. In these economies, monetary policy shocks command a constant price of risk but help to resolve economic understanding of additional salient features of the data, such as the term structure of yield volatility.

4.2 Data

This section describes the data set. The sample we study is at monthly frequency and runs from January 1990 to August 2012.
4.2.1 Survey data

We use survey forecasts from BlueChip Financial Forecasts Indicators (BCFF) to construct a new measure of monetary policy shocks. BCFF is a monthly publication providing extensive panel data on the expectations of professional economists working at leading financial institutions and service companies. Forecasted variables include Treasury yields and economic fundamentals. While the exact timing of the surveys is not published, the survey is usually conducted between the 25th and 27th of the month and mailed to subscribers within the first 5 days of the subsequent month. The resulting dataset represents an extensive and unique dataset to investigate the role of formation of expectations about monetary policy shocks.

The horizon of BCFFS forecasts ranges from the end of the current quarter to 5 quarters ahead (6 from January 1997). We obtain a set of constant maturity forecasts (from 1 to 4 quarters ahead), by interpolating linearly between adjacent horizons. Macroeconomic forecasts are expressed as annualized percentage changes between subsequent quarters: we obtain compound growth forecasts by chaining subsequent quarterly forecasts.

The resulting dataset of forecasts can be described as follows. Let $Z_t$ denote the time-$t$ realization of the economic or financial variable of interest, and let $E^n_{\cdot \mid \Omega_{n,t}}$ denote the expectation operator under the subjective measure of agent $n$ and conditional on her time-$t$ information set $\Omega_{n,t}$. The data manipulations described above allows us to obtain $Z_{n,t,h}^e$, the forecast of $Z_{t+h}$ made by agent $n$ at time $t$:

$$Z_{n,t,h}^e \overset{\Delta}{=} E^n \left[ Z_{t+h} \mid \Omega_{n,t} \right],$$

for quarterly horizons out to 1 year, $h = 3, 6, 9, 12$ months. Notice that this representation allows for incomplete information ($\Omega_{n,t}$), and difference in priors about the data-generating

$^6$Andrea Buraschi and Paul Whelan obtained the complete BCFF chapter archive directly from Wolters Kluwer and proceeded to enter manually the data. The digitization process required inputting around 750,000 entries of named forecasts plus quality control checking and was completed in a joint venture with the Federal Reserve Board.

$^7$For instance, suppose that as of April 2000, the 1Q- and 2Q-ahead GDP forecasts of agent $n$ are 5.00 and 6.00, respectively. This means that the agent expects GDP to increase by $(1 + \frac{5.00}{400})$ between April 2000 (the month of the forecast) and June 2000 (the end of current quarter), and by $(1 + \frac{6.00}{400})$ between end of June 2000 (the end of current quarter) and the end of September 2000 (the end of the next quarter). The (annualized) compound growth rate between April 2000 and September 2000 is obtained as $\left[ (1 + \frac{5.00}{400}) \cdot (1 + \frac{6.00}{400}) \right]^2$.
process (the expectation is taken under the subjective measure); the only assumption is that forecasts be rational in the sense of Muth (1961). We also construct consensus forecasts \( Z_{C,t,h}^e \), defined as the cross-sectional mean of the forecasts by all respondents at time \( t \):

\[
Z_{C,t,h}^e \equiv \frac{1}{N} \sum_{n=1}^{N} Z_{n,t,h}^e,
\]

where \( N \) denotes the size of the cross-section of forecasters.

The forecasts used here are real GDP (Real GNP until February 1992), Consumer Price Inflation, and the Federal Funds rate. Since real GDP, CPI, and federal funds rates are available at different frequencies (quarterly, monthly and daily, respectively), the quarterly values that the survey participants are asked to forecast are defined in different fashions. Let \( GDP_q(t) \), \( CPI_{m(t,j)} \), and \( FF_{d(t,j)} \) denote, respectively: (i) the seasonally adjusted value of real GDP at the end of the quarter that includes month \( t \); (ii) the seasonally adjusted value of CPI at the end of the \( j \)-th month of the quarter that includes month \( t \); and (iii) the value of the federal funds rate at the end of the \( j \)-th day of the quarter that includes month \( t \) (assumed to be 90, for simplicity). For each horizon \( h \), survey participants are asked to forecast are \( g_{q(t+h)} \), the quarter-over-prior-quarter percent change of seasonally-adjusted real GDP, expressed as an annualized rate:

\[
g_{q(t+h)} \equiv \left( \frac{GDP_{q(t+h)}}{GDP_{q(t+h-1)}} \right)^4 - 1;
\]

\( \pi_{q(t+h)} \), the quarter-over-prior-quarter percent change of the intra-quarter average of seasonally-adjusted CPI, expressed as an annualized rate:

\[
\pi_{q(t+h)} \equiv \left( \frac{1/3 \sum_{m=1}^{3} CPI_{m(t+h,j)}}{1/3 \sum_{j=1}^{3} CPI_{m(t+h-1,j)}} \right)^4 - 1;
\]

and, \( f_{q(t+h)} \), the average of intra-quarter daily federal funds rates:

\[
f_{q(t+h)} \equiv 1/90 \sum_{j=1}^{90} FF_{d(t+h,j)}.
\]

We denote the time-\( t \) forecasts of agent \( n \) for \( g_{q(t+h)} \), \( \pi_{q(t+h)} \), and \( f_{q(t+h)} \) by \( g_{n,t,h}^e \), \( \pi_{n,t,h}^e \), and \( f_{n,t,h}^e \), respectively. Figures 4.1 and 4.2 show the dynamics of the 1-year ahead consensus forecasts for the federal funds rate, inflation, and GDP growth.
Figure 4.1: Consensus federal funds forecasts

This figure plots the time series of $f_{C,t,12}^*$, the consensus 1-year ahead forecast for the federal funds rate.

We construct expected output gaps as follows. We obtain quarterly data on real GDP from the Bureau of Economic Analysis (BEA), and interpolate linearly to obtain monthly figures. We fit a Hodrick-Prescott filter (with a smoothing parameter of 14,400) to log output $y_t = \log(Y_t)$ and estimate the mean growth rate of the economy $g_t^*$ as the average log difference of output. We construct potential output $Y_t^*$ by taking the (exponential of the) trend component of the filtered series, and construct conditional estimates of future potential output (common across agents) as $E[Y_{t+h}^*|\Omega_t] = Y_t^* \exp(g_t^* \cdot h \cdot 3)$.\textsuperscript{8} Next, we obtain estimates

\textsuperscript{8}The construction of this expectations implicitly assumes that output is lognormally distributed and ignores a Jensen’s inequality term, which is quantitatively negligible.
Figure 4.2: Consensus macro forecasts

This figure plots the time series of $\pi_{C,t,12}^c$ and $g_{C,t,12}^c$, the consensus 1-year ahead forecast for inflation (black continuous line) and GDP growth (red dashed line).

of actual output using individual GDP growth forecasts, $E^n[Y_{t+h} | \Omega_{n,t}] = Y_t \cdot \left(1 + \frac{g_{n,t+1}^c}{400}\right) \cdot \left(1 + \frac{g_{n,t+2}^c}{400}\right) \ldots \left(1 + \frac{g_{n,t+h}^c}{400}\right)$. Finally, we construct the percentage projected output gap for horizon $h$ as $x_{n,t,h}^c = E^n[x_{t+h} | \Omega_{n,t}] = \left(\frac{E^n[Y_{t+h} | \Omega_{n,t}]}{E[Y_{t+h} | \Omega_{t}]} - 1\right) \cdot 100$. Since this definition of gap may suffer from look-ahead biases, we also construct real time output gaps by fitting the Hodrick-Prescott filter and estimating mean growth rates recursively over a 10-years look-backwards rolling window. The results are only marginally affected.
4.2.2 Quality of survey data

Survey data forecasts feature a number of advantages over forecasts implied by econometric approaches such as VARs (vector autoregressions). First, the specification of the VAR may not coincide with the data generating process in the mind of the agents. Second, agents may, contrary to the econometrician, observe a structural break in the sample of interest. Third, even if both the econometrician and the agents observe the data generating process, VAR forecasts still suffer from estimation error. Survey data allows obtaining direct measures of agents' expectations, dispensing with the need to posit and estimate a data generating process for the variable of interest.

We summarize the cross-sectional and time-series properties of BCFFS expectations by comparing their performance to an econometric benchmark. In particular, we first construct the sample equivalents of the forecasts for federal funds, real GDP growth rate, and CPI growth rate. Next, we assume that the quantities of interest $[\Delta f_t, y_t, \pi_t]'$ follow a first-order VAR. We fit the VAR recursively using a 25 years rolling window (100 quarterly observations), and use the estimated parameters to construct the benchmark forecasts. Since macro-economic data are released with a month lag, we always drop the last observations when estimating the VAR to ensure that forecasts are based on the actual real-time information set of agents. Finally we compare the forecasting errors of BCFFS versus VAR(1) expectations. Figure 4.3 summarizes the magnitude of BCFFS forecast errors (1 quarter horizon) relative to VAR forecast errors. The plots on the left represent the errors from a cross-sectional perspective: they show the time series of the number of agents in the cross-section whose forecast error is, in absolute value, less than the absolute value of the VAR forecast error. The plots on the right, on the other hand, summarize the forecasting ability of consensus (mean) forecasts: they show the difference between the absolute value of the VAR forecast error and the absolute value of the average BCFFS forecast error, so that a value above zero means that, on a specific quarter, the consensus forecasts performs better than the VAR forecast. The figure suggests that there is a strong time-series component in the ability of BCFFS surveys to beat VAR forecasts. Overall, the errors of consensus forecasts are, in absolute value, less than the forecast error of the VAR 85% (FF), 43% (GDP), and 65% (CPI) of the times.
Figure 4.3: The performance of BCFFS forecasts
The figure plots the time series of the number of agents in the cross-section whose forecast error is, in absolute value, less than the absolute value of the VAR forecast error (left panels), and the difference between the absolute value of the VAR forecast error and the absolute value of the average BCFFS forecast error (right panels). Forecast horizon: one quarter.

4.2.3 Macroeconomic activity data
We construct a proxy for the level of macroeconomic activity by following Ludvigson & Ng (2009) and Buraschi & Whelan (2012). Ludvigson & Ng (2009) find strong evidence linking bond returns to variations in the level of economic growth rate factors by running return predictability regressions on the principle components from a large panel of real, nominal,
and price-based variables. The identity and sources of the dataset are described in Ludvigson & Ng (2009); following Buraschi & Whelan (2012), we drop all price based information in order to interpret the resulting panel as a pure growth rate factor. Examples of price variables removed include: S&P dividend yield, the Federal Funds (FF) rate; 10 year T-bond; 10 year - FF term spread; Baa - FF default spread; and the dollar-Yen exchange rate. A small number of discontinued macro series are replaced with appropriate alternatives or dropped.\(^9\) We take the first principle component of the resulting dataset of 99 macro series as a proxy for the conditional mean of consumption growth, \(g_t\);\(^10\) Figure 4.4 plots \(g_t\), alongside the consensus GDP growth forecast for comparison purposes.

### 4.2.4 Bond data

We use Fama-Bliss data from CRSP of zero coupon bond prices (available at monthly frequency) with maturities between 1 and 5 years. The following notation is adopted. Define the date \(t\) log price of a \(n\)-year discount bond as \(p_t^{(n)}\). The yield of a bond is defined as \(y_t^{(n)} = -\frac{1}{n} p_t^{(n)}\). The date-\(t\) 1-year forward rate for the year from \(t + n - 1\) and \(t + n\) is \(f_t^{(n)} = p_t^{(n)} - p_t^{(n+1)}\). The log holding period return is the realised return on an \(n\)-year maturity bond bought at date \(t\) and sold as an \((n-1)\)-year maturity bond at date \(t+12\):

\[
    r_{t+12}^{(n)} = p_{t+12}^{(n-1)} - p_t^{(n)}.
\]

Excess holding period returns are denoted by:

\[
    x_{t+12}^{(n)} = r_{t+12}^{(n)} - y_t^{(1)}.
\]

Table 4.1 presents the summary statistics of bond excess returns for our sample, while Figure 4.5 plots their dynamics.

### 4.3 Monetary policy shocks

This section describes the methodology we use to construct monetary policy shocks. First, we provide a brief overview of the alternative approaches employed by the literature. Second,

\(^9\)Further details of the construction and macro series included are given in the appendix of Buraschi & Whelan (2012).

\(^{10}\) \(g_t\) explains around 90% of the unconditional variance of the panel of macroeconomic activity series.
Figure 4.4: Macroeconomic activity
This figure plots the time series of $g_t$ (black continuous line), the proxy for the level of macroeconomic activity described in the data section. For comparison purposes the figure also shows $g_{C,t,12}^c$, the consensus 1-year ahead forecast for GDP growth.

we introduce our identification scheme based on a Taylor rule and a panel of forecast data. Third, we report the statistical properties of the constructed series.

4.3.1 Overview of identification schemes
Since much of the decisions taken by the monetary authorities reflect non-monetary developments in the economy, the literature that studies the effect of monetary policy is typically concerned with policy shocks, rather than policy actions per se (Christiano et al. (1999)). A classic interpretation of monetary policy shocks is that they reflect changes in the prefer-
Table 4.1: Summary statistics of bond excess returns

This table presents the summary statistics of one-year bond excess returns constructed from log-prices of Fama-Bliss zero coupon bonds: \( r_{t+1}^{(n)} = \ln P_{t+12}^{(n)} - \ln P_t^{(n)} - y_t^{(1)} \). Bond maturities \( n \) range from 2 to 5 years. Sample period: 1990:1 - 2011:7 (last excess return is between 2011:7 and 2012:7).

<table>
<thead>
<tr>
<th>Maturity</th>
<th>2 Years</th>
<th>3 Years</th>
<th>4 Years</th>
<th>5 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
</tr>
<tr>
<td>Mean</td>
<td>0.94%</td>
<td>1.81%</td>
<td>2.58%</td>
<td>3.07%</td>
</tr>
<tr>
<td>Std Dev</td>
<td>1.34%</td>
<td>2.54%</td>
<td>3.56%</td>
<td>4.41%</td>
</tr>
<tr>
<td>Min</td>
<td>-2.37%</td>
<td>-5.25%</td>
<td>-6.90%</td>
<td>-8.39%</td>
</tr>
<tr>
<td>Max</td>
<td>3.63%</td>
<td>7.33%</td>
<td>10.32%</td>
<td>12.56%</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.11</td>
<td>-0.28</td>
<td>-0.37</td>
<td>-0.45</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.22</td>
<td>2.51</td>
<td>2.58</td>
<td>2.68</td>
</tr>
<tr>
<td>1st lag AC</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>2nd lag AC</td>
<td>0.88</td>
<td>0.86</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>3rd lag AC</td>
<td>0.80</td>
<td>0.77</td>
<td>0.74</td>
<td>0.71</td>
</tr>
</tbody>
</table>

In general, the strategies that seek to identify monetary policy shocks can be classified as being based on a policy rule, or not. The idea behind strategies based on policy rules is to impose as much structure on the feedback rule as needed to decompose policy actions into systematic and non-systematic components (policy shocks). In particular, researchers following this approach must make assumptions about the policy instrument, the functional form of the feedback rule, the arguments of the rule, and, importantly, the interaction of the arguments of the rule and monetary policy shocks. A common assumption, which underlies the entire literature on Taylor rules,\(^{11}\) is that monetary policy shocks are orthogonal to the arguments of the rule, so that they can be estimated from regression residuals. Christiano

\(^{11}\)The literature on Taylor rules is extensive: see, for instance, Taylor (1993), Clarida et al. (2000), Ang et al. (2007).
Figure 4.5: Bond excess returns
This figure plots one-year bond excess returns $r_{t+12}^{(n)}$ for bond maturities $n$ between 2 and 5 years. Sample period: 1990:1 - 2011:7 (last excess return is between 2011:7 and 2012:7).

*et al.* (1999) note that the economic content of this assumption is that the macroeconomic variables that the monetary policy authority looks at in setting policy are predetermined relative to the policy shocks: these variables do not respond to contemporaneous realizations of monetary policy shocks.

The second family of identification strategies, on the other hand, infers monetary policy shocks from data that suggests exogenous monetary policy actions. An example of this identification philosophy is Romer & Romer (1989), who use historical records to identify large monetary disturbances not caused by macroeconomic developments. Similarly, Romer & Romer (2004) use quantitative and narrative records to infer intended funds rates changes.
around FOMC meetings, and take into account the forecasts of the Federal Reserve to remove anticipatory components. An alternative approach, pioneered by Rudebusch (1998) and employed, among others, by Kuttner (2001), Cochrane & Piazzesi (2002), Gurkaynak et al. (2005), and Bernanke & Kuttner (2005), constructs monetary policy shocks from federal funds futures by exploiting their forward-looking nature. The advantage of this approach is that it overcomes the omitted-variables and time-varying parameter issues that Vector Autoregressions suffer from (Cochrane & Piazzesi (2002)).

4.3.2 Constructing a measure of path shocks

There are two types of monetary policy shocks: target shocks and path shocks. Target shocks represent exogenous variation in the conduct of monetary policy as reflected by the current behavior of the policy instrument (the target rate); all measures of monetary policy shocks introduced in the previous section belong to this category. Path shocks, on the other hand, capture exogenous variation in the projected path of monetary policy. Intuitively, path shocks reflect the surprises about future policy that can be inferred, for instance, from FOMC statements or interviews of members of the policy committee. Path shocks, unlike target shocks, cannot be easily identified via Taylor rule regressions because the FOMC does not reveal the projected evolution for the path of monetary policy;\textsuperscript{12} this is unfortunate, since path shocks represent the quantitatively most important source of variation in bond yields, especially for long maturities (Gurkaynak et al. (2005)). Our data on federal funds forecasts, however, allows us to measure the projected evolution for the path of monetary policy from the perspective of market participants; since we also observe market participants’ expectations about macroeconomic developments, we are able to identify path shocks, the exogenous variation in the path of monetary policy, by taking the residuals from a policy feedback rule estimated on federal funds and macro forecasts. This section describes the Taylor specification we adopt and discusses the restrictions it implies for forecast data.

\textsuperscript{12}Recently, there has been a change in the disclosure about future policy actions; see the last subsection for a discussion.
points. As standard in the literature, the output gap is defined as \( x_t = \left( \frac{Y_t}{Y_t^*} - 1 \right) \cdot 100 \), where \( Y_t \) and \( Y_t^* \) denote actual and potential output at time \( t \), respectively. Finally, let \( f_t \) denote the time \( t \) Federal Funds rate, with long run mean \( f \). The level of the federal funds rate can be decomposed into a systematic component \( f_t^* \) (the Taylor rule), and an orthogonal shock \( u_t \) (the target shock):

\[
f_t = f_t^* + u_t = \left( f + \beta (\pi_t - \pi) + \gamma x_t \right) + u_t,
\]

where \( u_t \perp f_t^* \). This characterization has a simple interpretation. The feedback rule \( f_t^* \) captures the systematic component of monetary policy. In the absence of disturbances to the economy and monetary policy shocks, the federal funds rate is constant and equal to \( f \). If output deviates from its potential level, or inflation from its target, the central bank intervenes to stabilize the economy: the parameters \( \beta \) and \( \gamma \) capture the sensitivity to inflation and output stabilization, respectively. The target shock \( u_t \), on the other hand, captures the non-systematic component of monetary policy: the orthogonality between \( u_t \) and the arguments of the Taylor rule means that it can be estimated as the residual of a simple time series regression of federal funds onto inflation and gap.

In practice, it has been observed that the central bank behaves less responsively to the state of the economy than implied by the benchmark Taylor rule, consistent with the idea that the central bank may have preferences over the degree of variability of federal funds rates. Also, the central bank may wish to respond to macro aggregates that are not realizing at time \( t \). In order to accommodate policy inertia and backward-/forward-looking policies, the benchmark rule can be extended to include lagged federal funds and lags/leads in its arguments, so that realized federal funds rates are described by:

\[
f_t = \rho(L)f_{t-1} + (1 - \rho)f_t^* + u_t = \rho(L)f_{t-1} + (1 - \rho)\left( f + \beta (\pi_{t+j} - \pi^*) + \gamma x_{t+k} \right) + u_t,
\]

where \( \rho(L) = \rho_1 + \rho_2 L + \ldots + \rho_m L^{m-1} \) and \( \rho = \rho(1) \) capture the degree of interest rate smoothing. Assuming that agents know the functional form and parameters of the policy
rule, while they can disagree about the future evolution of macroeconomic variables, the time-\(t\) expectation of \(f_{t+h}\) of agent \(n\) is given by:

\[
E^n[f_{t+h}|\Omega_{n,t}] = E^n[\rho(L)f_{t+h-1}|\Omega_{n,t}]
+ (1 - \rho) \left( f + \beta (E^n[\pi_{t+h+j}|\Omega_{n,t}] - \pi^*) + \gamma E^n[x_{t+h+k}|\Omega_{n,t}] \right)
+ E^n[u_{t+h}|\Omega_{n,t}],
\]

which, using the notation introduced in the data section, can be re-written as:

\[
f^e_{n,t,h} = \rho_1 f^e_{n,t,h-1} + \ldots + \rho_m f^e_{n,t,h-m}
+ (1 - \rho) \left( f + \beta (\pi^e_{n,t,h+j} - \pi^*) + \gamma x^e_{n,t,h+k} \right) + u^e_{n,t,h},
\]

or, in its consensus form, as:

\[
f^e_{C,t,h} = \rho_1 f^e_{C,t,h-1} + \ldots + \rho_m f^e_{C,t,h-m}
\underbrace{\left[ \text{Expected \hspace{1cm} inertia} \right]}_{\text{Expected \hspace{1cm} systematic \hspace{1cm} component}}
+ (1 - \rho) \left( f + \beta (\pi^e_{C,t,h+j} - \pi^*) + \gamma x^e_{C,t,h+k} \right) + u^e_{C,t,h}.
\]

This expression shows that the assumption that agents believe in a Taylor rule has two implications. First, it implies a restriction on the comovement of forecasts of federal funds, inflation, output gap, and monetary policy shocks. Second, it suggests that subjective expectations about future monetary policy shocks (path shocks) can be recovered from a panel of macroeconomic and financial forecasts.

### 4.3.3 Empirical features of the monetary policy shocks measure

All the specifications examined amount to choices for \(h\), the horizon of the federal funds forecasts, and \(i\), the horizon of the output gap and inflation forecasts, in the general model:

\[
f^e_{n,t,h} = \rho_1 f^e_{n,t,h-1} + \rho_2 f^e_{n,t,h-2} + (1 - \rho) \left( f + \beta (\pi^e_{n,t,i} - \pi^*) + \gamma x^e_{n,t,i} \right) + u^e_{n,t,h};
\]

all specifications include 2 smoothing terms, which are necessary to remove the persistent component of federal funds forecasts as in Clarida et al. (2000). We consider three families
of specifications based on the choice of the horizons of the forecasts for the federal funds and the arguments of the Taylor rule: 2 are contemporaneous, and 1 is forward looking. For each family, output gap is either constructed using full sample information (GAP 1), or recursively (GAP 2). This gives a total of 6 specifications, that are summarized in Table 4.2.

**Table 4.2: Taylor rule specifications**

This table describes different specifications of the general Taylor rule model

$$f_{n,t,h} = \rho_1 f_{n,t,h-1} + \rho_2 f_{n,t,h-2} + (1 - \rho) \left( f + \beta \left( \pi_{n,t,i}^e - \pi^* \right) + \gamma \pi_{n,t,i}^c \right) + u_{n,t,h}^e.$$  

The first row contains the horizon, in months, of the federal funds rate forecast. The second row contains the horizon, in months, of inflation and output gap forecasts. The third row describes the type of output gap employed; output gap 1 is constructed using full sample information, while output gap 2 is constructed recursively. The fourth and fifth rows report the test-statistic and p-value of the F-test for the joint significance of the individual effects. The final two rows report the test-statistic and p-value of the Hausman test for systematic differences between random and fixed effects coefficients (null hypothesis: random effects is appropriate).

<table>
<thead>
<tr>
<th>Spec</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>i</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>GAP</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F(FE)</td>
<td>7.69</td>
<td>7.56</td>
<td>8.86</td>
<td>8.85</td>
<td>8.19</td>
<td>8.31</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Hausman</td>
<td>52.91</td>
<td>50.61</td>
<td>32.71</td>
<td>37.17</td>
<td>31.28</td>
<td>36.23</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

These models could be estimated (i) using consensus data, via OLS; or, using panel data, via: (ii) pooled OLS (POLS); (iii) fixed effects (FE); (iv) random effects (RE). We estimate the model using fixed effects; the Appendix discusses the relationship among all estimation approaches. Table 4.2 reports the test-statistics and p-values for two tests: (i) an F-test for
the joint significance of individual effects, and (ii) the Hausman test for the null hypothesis that the difference between random and fixed coefficients is not systematic. In all cases, the F-statistic for the joint significance of agent dummies rejects the null of no significance; furthermore, the Hausman test always rejects the null of random effects. Taken together, these results indicate that fixed effects is preferable over estimation procedures based on consensus, pooled OLS, and random effects.

Given one of the 6 Taylor rule specifications, we construct a measure of policy path shocks, \( \text{PathShock}_t \), by taking the cross-sectional average of the estimated residuals:

\[
\text{PathShock}_t = \frac{1}{N} \sum_{n=1}^{N} \alpha_{nt,h}^e.
\]

Table 4.3 presents the summary statistics of the series, while Figure 4.6 plots their time series: the series feature a high level of comovement.

Figure 4.7 plots \( \text{PathShock} \) (Specification 1) and the level of macroeconomic activity \( g \), and highlights the NBER recessions in gray shade: with the exception of the last NBER recession, where both \( \text{PathShock} \) and \( g \) plunge, the two series feature little comovement, consistent with the economic assumption that monetary policy shocks are orthogonal to the macroeconomic environment.

It is natural to ask whether the monetary policy shocks extracted from macroeconomic surveys are also encoded by observable bond yields; this question is related to the general issue of whether yields can be inverted to reveal (span) the state vector. Duffee (2011) points out that the assumption of yield invertibility breaks down when the dimension of the state vector under the risk neutral measure is different from its dimension under the physical measure: intuitively, a factor may contemporaneously raise expected future short rates and decrease term premia, thus leaving the current term structure unaffected. Duffee (2011) suggests that one such factor may be economic growth: high growth states are likely to decrease risk aversion and increase the probability of future tightening. Therefore, to the extent that path shocks affect term premia and expectations about future short rates in the same direction, they should be recoverable from current yields. This argument, however, relies on the assumption that the path shocks measure we construct consists of policy path shocks expected by the marginal investor, which may not be the case. Hence, in general,
Table 4.3: Summary statistics of $\text{PathShock}$
This table shows the summary statistics of monetary policy path shocks $\text{PathShock}_t^i$, constructed as cross-sectional averages of the residuals from Taylor rules estimated over a panel of forecast data. Each series $(i)$ correspond to one of the 6 specifications described in Table 4.2. Sample period: 1990:1 - 2011:7.

<table>
<thead>
<tr>
<th>Spec</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
</tr>
<tr>
<td>Mean</td>
<td>0.39%</td>
<td>0.38%</td>
<td>0.35%</td>
<td>0.35%</td>
<td>0.37%</td>
<td>0.37%</td>
</tr>
<tr>
<td>Std Dev</td>
<td>4.61%</td>
<td>4.55%</td>
<td>5.58%</td>
<td>5.63%</td>
<td>5.51%</td>
<td>5.60%</td>
</tr>
<tr>
<td>Min</td>
<td>-10.59%</td>
<td>-10.68%</td>
<td>-10.60%</td>
<td>-11.04%</td>
<td>-11.72%</td>
<td>-12.33%</td>
</tr>
<tr>
<td>Max</td>
<td>14.21%</td>
<td>13.41%</td>
<td>14.80%</td>
<td>14.84%</td>
<td>14.69%</td>
<td>14.98%</td>
</tr>
<tr>
<td>Skew</td>
<td>0.10</td>
<td>-0.02</td>
<td>0.34</td>
<td>0.27</td>
<td>0.33</td>
<td>0.30</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.76</td>
<td>2.68</td>
<td>2.37</td>
<td>2.39</td>
<td>2.44</td>
<td>2.46</td>
</tr>
<tr>
<td>1st lag AC</td>
<td>0.80</td>
<td>0.79</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>2nd lag AC</td>
<td>0.67</td>
<td>0.66</td>
<td>0.73</td>
<td>0.73</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>3rd lag AC</td>
<td>0.59</td>
<td>0.57</td>
<td>0.62</td>
<td>0.62</td>
<td>0.61</td>
<td>0.62</td>
</tr>
</tbody>
</table>

the ability of yields to span path shocks will largely depend on the relative difference in dimensions of the state vector under the physical versus BCFFS-consensus measure.

We assess the ability of yields to span path shocks via linear regressions of monetary policy shocks on a constant and information contained in the yield curve:

$$\text{PathShock}_t = \text{const.} + Z_t'\beta + \epsilon_t.$$  
We consider three proxies for yield curve information $Z_t$: (i) the first three principal components of yields ($PC1$, $PC2$, and $PC3$), (ii) the factor of Cochrane & Piazzesi (2005) ($CP$), and (iii) the two volatility factors constructed by Le & Singleton (2013) ($LS1$ and $LS2$). Table 4.4 presents the results of the regressions. There are three conclusions to be drawn; all are robust to the choice of the specification for $\text{PathShock}$. First, the only statistically significant loading in (i) is $PC2$ (t-stats above 2.2 for all specifications), which can be interpreted
Figure 4.6: PathShock
This figure plots monetary policy path shocks PathShock, constructed as cross-sectional
averages of the residuals from Taylor rules estimated over a panel of forecast data. Each
series corresponds to one of the 6 specifications described in Table 4.2. Sample period:

as a slope factor: path shocks are high when the term structure is steep. Second, the comove-
ment of path shocks and the Cochrane & Piazzesi (2005) factor is positive and statistically
significant (t-stats above 2.2 for all specifications): given the ability of the Cochrane & Pi-
azzesi (2005) factor to predict excess returns, this finding suggests that also PathShock may
capture the time variation of risk premia. Third, path shocks load on LS2 in a statistically
significant way (t-stats above 2.8 for all specifications). Hence, to a large extent, PathShock
seem to be spanned by yield curve information. Figure 4.8 confirms these results visually by
Figure 4.7: Pro cyclicality of $\mathcal{P}athShock$

This figure plots monetary policy path shocks $\mathcal{P}athShock$ (specification 1), and macroeconomic activity, $g$. Areas shaded in gray indicate NBER recessions. Sample period: 1990:1 - 2011:7.

plotting $\mathcal{P}athShock$ (specification 1), alongside the yield curve information factors that have the highest explanatory power for its dynamics: $PC2$, $CP$, and $LS2$. 
Table 4.4: Spanning

The table reports the output from regressions of \( \text{PathShock} \) on a constant and information contained in the yield curve: \( \text{PathShock}_t = \text{const.} + Z_t' \beta + \epsilon_t \). The proxies for yield curve information \( Z_t \) are the first three principal components of yields (\( PC1, PC2, \) and \( PC3 \)), the factor of Cochrane & Piazzesi (2005) (\( CP \)), and the two volatility factors constructed by Le & Singleton (2013) (\( LS1 \) and \( LS2 \)). Each panel reports the results for one of the 6 proxies of expected monetary policy described by Table 4.2. T-statistics, reported below the point estimates, are corrected for auto-correlation and heteroskedasticity using Newey-West errors (18 lags). \( \bar{R}^2 \) is the adjusted \( R^2 \). Both left and right hand variables are standardized. A constant is included but not reported. Sample period: 1990:1 - 2007:12.

<table>
<thead>
<tr>
<th></th>
<th>( PC1 )</th>
<th>( PC2 )</th>
<th>( PC3 )</th>
<th>( CP )</th>
<th>( LS1 )</th>
<th>( LS2 )</th>
<th>( \bar{R}^2 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spec. 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.04</td>
<td>0.32</td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13.57%</td>
<td>8.22</td>
</tr>
<tr>
<td>-0.31</td>
<td>2.29</td>
<td>-1.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td>12.63%</td>
<td>7.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
<td>0.54</td>
<td>24.13%</td>
<td></td>
<td></td>
<td>13.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.44</td>
<td>2.86</td>
<td>24.13%</td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Spec. 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.04</td>
<td>0.31</td>
<td>-0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.38%</td>
<td>8.07</td>
</tr>
<tr>
<td>-0.33</td>
<td>2.33</td>
<td>-1.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td>12.32%</td>
<td>8.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.52</td>
<td>23.02%</td>
<td></td>
<td></td>
<td>15.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.30</td>
<td>2.78</td>
<td>23.02%</td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

Continued on next page
Continued from previous page

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>CP</th>
<th>LS1</th>
<th>LS2</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.01</td>
<td>0.41</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td>15.81%</td>
<td>13.29</td>
</tr>
<tr>
<td>-0.09</td>
<td>3.06</td>
<td>-0.41</td>
<td>0.46</td>
<td></td>
<td></td>
<td>21.06%</td>
<td>13.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.68</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.96</td>
<td>0.60</td>
<td>31.33%</td>
<td>26.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td>4.21</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Specification 4

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>CP</th>
<th>LS1</th>
<th>LS2</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.01</td>
<td>0.40</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
<td>15.20%</td>
<td>15.81</td>
</tr>
<tr>
<td>-0.10</td>
<td>3.09</td>
<td>-0.47</td>
<td>0.46</td>
<td></td>
<td></td>
<td>21.00%</td>
<td>15.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.95</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.94</td>
<td>0.59</td>
<td>31.93%</td>
<td>28.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.24</td>
<td>4.39</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Specification 5

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>CP</th>
<th>LS1</th>
<th>LS2</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.02</td>
<td>0.38</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
<td>13.52%</td>
<td>10.44</td>
</tr>
<tr>
<td>-0.18</td>
<td>2.80</td>
<td>-0.22</td>
<td>0.44</td>
<td></td>
<td></td>
<td>18.76%</td>
<td>11.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.37</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.56</td>
<td>28.85%</td>
<td>22.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
<td>3.76</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Specification 6

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>CP</th>
<th>LS1</th>
<th>LS2</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.02</td>
<td>0.37</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
<td>13.28%</td>
<td>13.27</td>
</tr>
<tr>
<td>-0.18</td>
<td>2.88</td>
<td>-0.31</td>
<td>0.44</td>
<td></td>
<td></td>
<td>19.23%</td>
<td>13.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.70</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.56</td>
<td>30.62%</td>
<td>25.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.02</td>
<td>4.10</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
4.4 Bond return predictability

This section establishes the key empirical results of the chapter: path shocks drive the time variation of bond risk premia. We obtain these results via classic return predictability projections: we regress one-year holding period bond excess returns on lagged PathShock, and test for the statistical and economic significance of the slope coefficient.

We verify the robustness of the results from a variety of perspectives: specifications of PathShock, bond maturities, and subsamples. First, we conduct the analysis for each of the 6
specifications of \textit{PathShock} introduced above. Second, we assess the statistical and economic significance of the slope coefficients in predictive regressions for bonds with maturities ranging between 2 and 5 years. Third, we analyze the stability of the coefficients in a subsample that excludes the crisis. Fourth, we conduct an out of sample test and check whether \textit{PathShock} is also priced in the cross-section of equity returns.

\subsection*{4.4.1 Univariate regressions}

Table 4.5 reports the estimation output of regressions:

\[ r x_{t+12}^{(n)} = \text{const.} + \beta_{PS}^{(n)} \text{PathShock}_t + \epsilon_{t+12}^{(n)}. \]

Each panel reports the results for one of the specifications of \textit{PathShock} described by Table 4.2; bond maturities \( n \) range from 2 to 5 years. The estimates on the left are for the entire sample (the last excess return is defined between 2011:7 and 2012: 7), while those on the right exclude the financial crisis (the last excess return is defined between 2007:6 and 2008: 6). Reported \( R^2 \) are adjusted, and all t-statistics employ Newey-West standard errors (18 lags); all left and right hand side variables are standardized, so that the coefficients can be interpreted as standard deviation changes of the regressand for a unit standard deviation change in the regressor. For illustrative purposes, Figure 4.9 plots the time series of excess returns of bonds with 2 and 5 years maturities, alongside the time series of \textit{PathShock} (specification 1).
Figure 4.9: PathShock and excess returns
This figure plots expected monetary policy path shocks PathShock (specification 1) and excess returns on 2 and 5 years bonds. Sample period: 1990:1 - 2011:7.

In all projections, PathShock explains the time variation in bond excess returns in an economically and statistically significant fashion. Estimated slope coefficients are between 0.35 and 0.43: a one standard deviation increase in path shocks predicts, on average, an increase of fitted excess returns by 40% standard deviations. Adjusted $R^2$ range between 12% and 18%, suggesting that besides its predictive ability, monetary policy path shocks also explain a large fraction of the overall variation of realized excess returns. Also the statistical significance is large: all t-statistics reject the null hypothesis of no predictability for standard levels of economic significance. Besides a slight decline in maturity of the economic magnitude of the slope coefficients, there are no noticeable patterns across specifications of the right hand side variable and sample period. Overall, the results suggest that path shocks
are not only a statistically and economically significant, but also robust predictor of future realized excess returns.

4.4.2 Bivariate regressions

Since levels of macroeconomic activity predict bond returns (see, for instance, Ludvigson & Ng (2009)), it is important to ensure that our measure of monetary policy shocks does not in fact proxy for underlying real activity factors: while monetary policy shocks are, by construction, orthogonal to the current macroeconomic developments, the same is not true, in general, for their expected counterparts. In order to control for the level of macroeconomic activity, we run bivariate predictive regressions:

\[ r_{t+12}^{(n)} = const. + \beta_{PS}^{(n)} PathShock_t + \beta_g^{(n)} g_t + \epsilon_{t+12}^{(n)} \]

Table 4.6 shows the results. The loadings on \( g \) are negative and significant in all specifications; both the statistical and economic significance are strongest for \( n = 2 \) and decrease in maturity. The inclusion of \( g \) in the set of regressors leads to a sizable increase in the adjusted \( R^2 \) relative to the univariate case, especially for short maturities and in the 1990-2007 sample. In the case of bonds with 2 years maturities, the average increase in \( R^2 \) is between 10\% and 15\% in the 1990-2007 period, and between 5\% and 10\% for the full sample; the effect decreases in maturity, becoming negligible for 5 years bonds. Nevertheless, the economic and statistical significance of expected monetary policy path shocks is only barely affected.
4.4.3 Predictability in the financial crisis

Tables 4.5 and 4.6 document that the economic and statistical significance of predictability is largely unaffected by the inclusion of the last financial crisis in the sample. This may sound, at first sight, a little surprising. Our predictor, \textit{PathShock}, is based on the notion that the federal fund rate is the instrument of monetary policy, a tool that has lost its flexibility and effectiveness in the context of the ZLB (zero lower bound) characterizing the US monetary landscape since the end of 2008.\textsuperscript{13} Unable to cut federal funds targets any further, US monetary authorities have started considering forward guidance\textsuperscript{14} and QE (Quantitative Easing)\textsuperscript{15} as alternative policy instruments (see Woodford (2012) for an extensive discussion).

Given such profound changes in the way that monetary policy is implemented, are measures of monetary policy shocks based on Taylor rule residuals appropriate and, more specifically, is \textit{PathShock} suitable to measure exogenous variation in monetary policy? \textit{Target} shocks, measured as residuals from Taylor rules estimated on \textit{current} federal funds, are indeed meaningless for two reasons. First, Taylor rules imply negative nominal federal funds rates in negative GDP growth and low inflation scenarios, thus ceasing to be adequate representations of the systematic and exogenous components of monetary policy. Second, surprises about current federal funds targets are little informative about how monetary policy is actually conducted in practice, since the Federal Reserve Bank has, de facto, switched its policy instrument from current federal funds targets to forward guidance (at least temporarily).

\textsuperscript{13}During the first turmoil and Lehmans’ collapse, the Fed engaged in a particularly intense series of target rate cuts: between 18 September 2007 and 16 December 2008, the target fed funds rate was decreased on each of the 10 FOMC meetings, going from 4.75% down to a 0%-0.25% range. Between 16 December 2008 and December 2012, the Fed maintained the target rate in the 0%-0.25% range uninterruptedly.

\textsuperscript{14}Forward guidance represents the result of a decade long process of changes in the strategy underpinning policy communication. The structure of FOMC statements has been modified to include: (i) an economic outlook, in January 2000; (ii) qualitative statements about future policy inclinations, in August 2003; (iii) calendar-based guidance, in August 2011; (iv) outcome-based guidance, in December 2012.

\textsuperscript{15}Quantitative Easing policies consist of purchases, by the central bank, of specified quantities of long term financial assets. Our sample includes two instances of QE policies: (i) QE1, between late 2008 and 2009; and (ii) QE2, between the second quarters of 2010 and 2011. While QE1 consisted of purchases of MBS, Treasuries, and Agency securities, QE2 focused only on the purchase on long term Treasury securities. See Gagnon \textit{et al.} (2010) and Krishnamurthy \& Vissing-Jorgensen (2010) for further details about QE policies and their quantitative impact on financial securities.
These criticisms, however, do not apply to residuals from Taylor rules estimated over expected future federal funds, and therefore to PathShock. First, despite the 0%-0.25% range imposed by the Fed onto current federal funds rates since December 2008, expected federal funds rates have featured noticeable volatility over the same period (see Figure 4.1). Second, expected future short rates are precisely the instrument of policies based on forward guidance: being a residual from a Taylor rule estimated on expectations of future short rates, PathShock is, by construction, a measure of the exogenous variation in forward guidance. As a consequence, PathShock is particularly suitable to measure exogenous monetary policy shocks in the recent monetary environment.

There is no clear interpretation of PathShock in terms of QE policies; nevertheless, it is interesting to study the comovement between forward guidance and QE policies from the perspective of BCFFS survey participants. To this end, we construct measures of expected QE path shocks by looking at the change in consensus yield spreads between 1 and 4 quarter horizons.\textsuperscript{16} We consider four measures of QE path shocks, each based on a specific yield spread survey forecast: (i) TS5Y, TS10Y, TS30Y: consensus increase in the spread between the 5-, 10-, or 30-years Treasury yield and the federal funds rate; (ii) MTGS: consensus increase in the spread between the mortgage yield and the 30-years Treasury yield. We consider measures based on both Treasury and mortgage securities because the Fed has intervened in both markets during QE1 and QE2. All these measures answer the question, by how much do agents expect that Treasury or mortgage spreads will rise over the next 4 quarters? Table 4.4.3 shows the pairwise correlations between PathShock and the four measures of QE path shocks in the full, pre-QE, and during-QE samples, taking November 2008 as the start date of the QE sample. The Table contains three interesting results. First, the correlation of PathShock with Treasury-based measures of QE is negative in all subsamples: when agents expect the Fed to tighten (high values of PathShock), they also expect the wedge between long and short rates to decline (low values of TS5Y, TS10Y, TS30Y). Second, the correlation of PathShock with Mortgage-based measures of QE is positive in all

\textsuperscript{16}As discussed earlier, any policy may be decomposed into systematic and exogenous components. Our measures of QE, however, are not orthogonalised. In fact, there is not qualitative evidence that the Federal reserve follows a rule in implementing QE policies, and no rule could be estimated quantitatively in such a short sample (QE policies started in 2008).
subsamples: an expansionary forward guidance stance (low values of \textit{PathShock}) is, in the mind of agents, expected to be coordinated with QE policies that aim to reduce the premium on mortgage securities (low values of $MTGS$). Third, all correlations rise, in absolute value, in the QE-subsample, suggesting that the switch to forward guidance policies has increased the influence of the communication of the Federal Reserve onto agents’ expectations.
Table 4.5: Bond return predictability: $PathShock$

The table reports the output from regressions of annual bond excess returns on a constant and expected monetary policy shocks: $r_{x_{t+12}} = const. + \beta_{PS}^{(n)}PathShock_t + \epsilon_{t+12}^{(n)}$. Bond maturities ($n$) range from 2 to 5 years. Each panel reports the results for one of the 6 proxies of expected monetary policy described by Table 4.2. The left panels report the results for the full sample (the last observation is the excess return that realized between 2011:7 and 2012:7), while the right panels report the results for the sample excluding the crisis (the last observation is the excess return that realized between 2007:6 and 2008:6). T-statistics, reported below the point estimates, are corrected for auto-correlation and heteroskedasticity using Newey-West errors (18 lags). $\bar{R}^2$ is the adjusted $R^2$. Both left and right hand variables are standardized. A constant is included but not reported.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$PathShock$</td>
<td>$\bar{R}^2$</td>
<td>$PathShock$</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>17.88%</td>
<td>18.18%</td>
</tr>
<tr>
<td></td>
<td>3.88</td>
<td>3.61</td>
</tr>
<tr>
<td>3</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>16.93%</td>
<td>16.84%</td>
</tr>
<tr>
<td></td>
<td>3.85</td>
<td>3.48</td>
</tr>
<tr>
<td>4</td>
<td>0.41</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>16.18%</td>
<td>15.62%</td>
</tr>
<tr>
<td></td>
<td>3.74</td>
<td>3.30</td>
</tr>
<tr>
<td>5</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>13.35%</td>
<td>13.37%</td>
</tr>
<tr>
<td></td>
<td>3.40</td>
<td>3.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$PathShock$</td>
<td>$\bar{R}^2$</td>
<td>$PathShock$</td>
</tr>
<tr>
<td>2</td>
<td>0.38</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>14.00%</td>
<td>15.53%</td>
</tr>
<tr>
<td></td>
<td>3.13</td>
<td>3.07</td>
</tr>
<tr>
<td>3</td>
<td>0.37</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>13.73%</td>
<td>14.62%</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>3.00</td>
</tr>
<tr>
<td>4</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>14.05%</td>
<td>13.70%</td>
</tr>
<tr>
<td></td>
<td>3.35</td>
<td>2.89</td>
</tr>
<tr>
<td>5</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>12.05%</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>2.77</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.37</td>
<td>13.56%</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td></td>
<td>3.03</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>12.83%</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>2.92</td>
<td></td>
<td>3.05</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>11.81%</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>2.78</td>
<td></td>
<td>3.03</td>
</tr>
<tr>
<td>5</td>
<td>0.31</td>
<td>9.38%</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td></td>
<td>2.95</td>
</tr>
</tbody>
</table>

**Specification 4**

<table>
<thead>
<tr>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.36</td>
<td>12.54%</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>2.76</td>
<td></td>
<td>2.80</td>
</tr>
<tr>
<td>3</td>
<td>0.35</td>
<td>12.10%</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>2.76</td>
<td></td>
<td>2.84</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>11.72%</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>2.79</td>
<td></td>
<td>2.86</td>
</tr>
<tr>
<td>5</td>
<td>0.32</td>
<td>9.69%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.62</td>
<td></td>
<td>2.82</td>
</tr>
</tbody>
</table>

**Specification 5**

<table>
<thead>
<tr>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.35</td>
<td>11.95%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.71</td>
<td></td>
<td>2.68</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>11.36%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.64</td>
<td></td>
<td>2.69</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>10.30%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td></td>
<td>2.67</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
<td>8.01%</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>2.22</td>
<td></td>
<td>2.57</td>
</tr>
</tbody>
</table>

**Specification 6**

<table>
<thead>
<tr>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
<th>PathShock</th>
<th>$\hat{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.35</td>
<td>11.64%</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>2.63</td>
<td></td>
<td>2.55</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>11.20%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.61</td>
<td></td>
<td>2.58</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>10.82%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2.61</td>
<td></td>
<td>2.59</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
<td>8.84%</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>2.42</td>
<td></td>
<td>2.55</td>
</tr>
</tbody>
</table>
Table 4.6: Bond return predictability: $\mathcal{P}_{\text{PathShock}}$ and $g$

The table reports the output from regressions of annual bond excess returns on a constant, expected monetary policy shocks, and levels of macroeconomic activity: $r_{t+12} = \text{const.} + \beta_{PS}^{(n)} \mathcal{P}_{\text{PathShock}} + \beta_{g}^{(n)} g_t + \epsilon_{t+12}^{(n)}$. Bond maturities ($n$) range from 2 to 5 years. Each panel reports the results for one of the 6 proxies of expected monetary policy described by Table 4.2. The left panels report the results for the full sample (the last observation is the excess return that realized between 2011:7 and 2012:7), while the right panels report the results for the sample excluding the crisis (the last observation is the excess return that realized between 2007:6 and 2008:6). T-statistics, reported below the point estimates, are corrected for autocorrelation and heteroskedasticity using Newey-West errors (18 lags). $\bar{R}^2$ is the adjusted $R^2$. Both left and right hand variables are standardized. A constant is included but not reported.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mathcal{P}_{\text{PathShock}}$</td>
<td>$g$</td>
<td>$\bar{R}^2$</td>
<td>$\mathcal{P}_{\text{PathShock}}$</td>
<td>$g$</td>
<td>$\bar{R}^2$</td>
</tr>
<tr>
<td><strong>Specification 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.37</td>
<td>-0.25</td>
<td>23.70%</td>
<td>0.25</td>
<td>-0.37</td>
<td>23.52%</td>
</tr>
<tr>
<td></td>
<td>3.55</td>
<td>-2.14</td>
<td></td>
<td>2.08</td>
<td>-3.96</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>-0.23</td>
<td>21.59%</td>
<td>0.27</td>
<td>-0.29</td>
<td>23.08%</td>
</tr>
<tr>
<td></td>
<td>3.44</td>
<td>-2.23</td>
<td></td>
<td>2.17</td>
<td>-3.26</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.36</td>
<td>-0.18</td>
<td>18.86%</td>
<td>0.29</td>
<td>-0.23</td>
<td>19.37%</td>
</tr>
<tr>
<td></td>
<td>3.28</td>
<td>-1.72</td>
<td></td>
<td>2.21</td>
<td>-2.67</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>-0.16</td>
<td>15.45%</td>
<td>0.29</td>
<td>-0.16</td>
<td>14.96%</td>
</tr>
<tr>
<td></td>
<td>2.99</td>
<td>-1.78</td>
<td></td>
<td>2.25</td>
<td>-1.98</td>
<td></td>
</tr>
<tr>
<td><strong>Specification 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>-0.31</td>
<td>22.97%</td>
<td>0.23</td>
<td>-0.40</td>
<td>28.04%</td>
</tr>
<tr>
<td></td>
<td>3.32</td>
<td>-2.70</td>
<td></td>
<td>1.85</td>
<td>-4.24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.35</td>
<td>-0.28</td>
<td>21.16%</td>
<td>0.25</td>
<td>-0.32</td>
<td>22.54%</td>
</tr>
<tr>
<td></td>
<td>3.30</td>
<td>-2.89</td>
<td></td>
<td>1.95</td>
<td>-3.57</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.36</td>
<td>-0.23</td>
<td>18.89%</td>
<td>0.26</td>
<td>-0.26</td>
<td>18.78%</td>
</tr>
<tr>
<td></td>
<td>3.27</td>
<td>-2.41</td>
<td></td>
<td>2.00</td>
<td>-3.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>-0.20</td>
<td>15.90%</td>
<td>0.27</td>
<td>-0.19</td>
<td>14.45%</td>
</tr>
<tr>
<td></td>
<td>3.10</td>
<td>-2.50</td>
<td></td>
<td>2.08</td>
<td>-2.32</td>
<td></td>
</tr>
</tbody>
</table>

*Continued on next page*
Continued from previous page

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{PathShock}$</td>
<td>$g$</td>
</tr>
<tr>
<td><strong>Specification 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.33</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>2.91</td>
<td>-2.13</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>2.82</td>
<td>-2.23</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td>2.65</td>
<td>-1.82</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>2.34</td>
<td>-1.87</td>
</tr>
<tr>
<td><strong>Specification 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>3.09</td>
<td>-2.51</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>3.03</td>
<td>-2.65</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>2.94</td>
<td>-2.22</td>
</tr>
<tr>
<td>5</td>
<td>0.31</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>2.70</td>
<td>-2.28</td>
</tr>
<tr>
<td><strong>Specification 5</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.31</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>2.57</td>
<td>-2.06</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>2.48</td>
<td>-2.14</td>
</tr>
<tr>
<td>4</td>
<td>0.29</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>-1.75</td>
</tr>
<tr>
<td>5</td>
<td>0.26</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>2.04</td>
<td>-1.79</td>
</tr>
<tr>
<td><strong>Specification 6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.33</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>2.85</td>
<td>-2.43</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>2.79</td>
<td>-2.56</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>2.70</td>
<td>-2.14</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>2.46</td>
<td>-2.20</td>
</tr>
</tbody>
</table>

107
<table>
<thead>
<tr>
<th>Sample period</th>
<th>TS5y</th>
<th>TS10y</th>
<th>TS30y</th>
<th>MTGS</th>
<th>nobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>-0.21</td>
<td>-0.28</td>
<td>-0.31</td>
<td>0.20</td>
<td>264</td>
</tr>
<tr>
<td>Pre-QE</td>
<td>-0.06</td>
<td>-0.14</td>
<td>-0.19</td>
<td>0.09</td>
<td>226</td>
</tr>
<tr>
<td>During QE</td>
<td>-0.83</td>
<td>-0.91</td>
<td>-0.89</td>
<td>0.43</td>
<td>38</td>
</tr>
</tbody>
</table>

**Table 4.7: PathShocks and QE path shocks**

The table reports the correlation of *PathShocks* (Specification 1) with four measures of QE path shocks: *TS5y, TS10y, TS30y,* and *MTGS*. The first, second, and third rows report the pairwise correlations for the full, pre-QE (up until November 2008), and QE sample (after November 2008), respectively.

### 4.4.4 Monetary policy shocks and equity markets

We conduct an out-of-sample test of our measure of monetary policy path shocks by asking whether it is priced in the cross-section of equity returns. The effect of monetary policy on real variables such as corporate earnings is likely to be unequal across horizons and industries. Companies may feature different sensitivities towards central bank intervention owing to, for instance, cross-sectional variation in cash flow duration: if monetary policy affects only short term GDP growth, for instance, growth companies will be the least affected by tightening cycles. To the extent that investors cannot diversify away exposure to central bank intervention, monetary policy shocks may be a priced risk factor in the cross-section of equity returns.\(^{17}\)

We follow two steps to test whether monetary policy shocks are priced in the cross-section of equity returns. First, we construct a portfolio that mimics monetary policy path shocks. Second, we ask whether exposure to path shocks is marginally priced by means of cross-sectional asset pricing tests. Throughout, we use Fama & French (1993) 100 value and size portfolios (available on Kenneth French's website) as test assets.

---

\(^{17}\)See Palomo & Li (2010) for a general equilibrium model where monetary policy is a priced risk factor in the cross-section of equity returns.
In order to construct a portfolio that mimics monetary policy path shocks, we first run time-series regressions of the excess returns of test portfolios \(RX^i_t, i = 1, \ldots, 100\) on \(\text{PathShock}\):

\[RX^i_t = \text{const.} + \beta^i_{\text{ps}} \text{PathShock}_t + \epsilon^i_t.\]

Next, we sort the 100 portfolios into ten deciles based on their sensitivity to monetary policy shocks \(\beta^i_{\text{ps}}\), and construct the factor mimicking portfolio \(m_{ps_t}\) as the average of the excess returns of the top decile (\(\text{Top}\)) portfolios, minus the average of the excess returns of the bottom decile (\(\text{Bottom}\)) portfolios:

\[m_{ps_t} \triangleq \frac{1}{10} \sum_{i \in \text{Top}} RX^i_t - \frac{1}{10} \sum_{i \in \text{Bottom}} RX^i_t.\]

Figure 4.10 shows the value and size characteristics of the portfolios used to construct \(m_{ps_t}\) (\(\text{PathShock}\) defined by specification 1).

Next, we run standard cross-sectional, two stage, asset pricing tests. In particular, we test whether the risk premium associated with exposure to \(m_{ps_t}\) is significant, in an economically and statistically way, over and beyond the premia associated with Fama & French (1993) factors: \(mkt_t, smb_t,\) and \(hml_t\). The first stage of the asset pricing test requires running time-series regressions of test assets’ excess returns on candidate risk factors to estimate factor exposures \(\beta\):

\[RX^i_t = \alpha^i + \beta^i_{\text{mkt}} mkt_t + \beta^i_{\text{smb}} smb_t + \beta^i_{\text{hml}} hml_t + \beta^i_{\text{mps}} m_{ps_t} + \epsilon^i_t.\]

In the second stage, the prices of risk \(\lambda\) associated with factor exposures \(\beta\) are estimated via cross-sectional regressions of average excess returns on factor exposures:

\[\overline{RX}^i = \text{const.} + \lambda_{\text{mkt}} \beta^i_{\text{mkt}} + \lambda_{\text{smb}} \beta^i_{\text{smb}} + \lambda_{\text{hml}} \beta^i_{\text{hml}} + \lambda_{\text{mps}} \beta^i_{\text{mps}} + \epsilon^i.\]

The economic and statistical significance of the prices of risk can be read off the magnitude and t-statistics of the \(\lambda\) estimates. Besides standard Newey-West t-statistics, we also calculate Shanken (1992) t-statistics to account for the fact that the regressors are estimated with error in the first-stage regressions.

Table 4.8 reports the (annualized) risk premia estimates and t-statistics for our four candidate pricing factors \(mkt_t, smb_t, hml_t, m_{ps_t}\); each specification constructs \(m_{ps_t}\) using
one of the six specification of PathShock described above. The economic significance is sizable: ranging from 3.63% to 5.39% on an annualized basis, the risk premium earned in compensation for holding monetary policy risk is second only to the market risk premium (the only exception is specification 1). Also the statistical significance is large: with the exception of specification 1, the null hypothesis $H_0 : \lambda_{mps} = 0$ is always rejected at standard levels of significance. Overall, the message of Table 4.8 is that path shocks are an important source of priced risk not only in bond markets, but also for the cross-section of stock returns.
Table 4.8: Monetary policy shocks and equity returns

The table reports risk premium estimates ($\lambda$) for the 4-factor equity asset pricing model $E[RX^t] = \beta^\mu \lambda$. The candidate risk factors are the market excess return ($mkt$), Fama & French (1993) value and size factors ($smb$ and $hml$), and the portfolio mimicking monetary policy shocks ($mps$): $\beta^\mu = [\beta^\mu_{mkt} \ \beta^\mu_{smb} \ \beta^\mu_{hml} \ \beta^\mu_{mps}]'$ and $\lambda' = [\lambda_{mkt} \ \lambda_{smb} \ \lambda_{hml} \ \lambda_{mps}]'$. Each specification uses $mps$ constructed from one of the 6 proxies of expected monetary policy shocks described in Table 4.2. Factor betas are estimated in first-stage time series regressions via OLS. For each specification: the first row reports (annualized) risk premia estimates; the second row reports t-statistics corrected for auto-correlation and heteroskedasticity using Newey-West errors (18 lags); the third row reports t-statistics that employ Shanken (1992) correction. Sample period: 1990:1 - 2011:7.

<table>
<thead>
<tr>
<th>Specification</th>
<th>mkt</th>
<th>smb</th>
<th>hml</th>
<th>mps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.17%</td>
<td>2.21%</td>
<td>4.03%</td>
<td>3.63%</td>
</tr>
<tr>
<td></td>
<td>7.91</td>
<td>3.89</td>
<td>2.92</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td>1.83</td>
<td>0.83</td>
<td>1.60</td>
<td>1.45</td>
</tr>
<tr>
<td>2</td>
<td>5.99%</td>
<td>2.52%</td>
<td>3.79%</td>
<td>4.40%</td>
</tr>
<tr>
<td></td>
<td>8.69</td>
<td>4.52</td>
<td>2.90</td>
<td>5.03</td>
</tr>
<tr>
<td></td>
<td>1.77</td>
<td>0.95</td>
<td>1.50</td>
<td>2.17</td>
</tr>
<tr>
<td>3</td>
<td>6.43%</td>
<td>2.12%</td>
<td>4.16%</td>
<td>4.40%</td>
</tr>
<tr>
<td></td>
<td>8.67</td>
<td>3.55</td>
<td>3.28</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>1.91</td>
<td>0.80</td>
<td>1.65</td>
<td>2.42</td>
</tr>
<tr>
<td>4</td>
<td>6.41%</td>
<td>2.31%</td>
<td>3.89%</td>
<td>5.39%</td>
</tr>
<tr>
<td></td>
<td>10.79</td>
<td>4.62</td>
<td>3.57</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>1.89</td>
<td>0.87</td>
<td>1.54</td>
<td>2.68</td>
</tr>
<tr>
<td>5</td>
<td>6.27%</td>
<td>2.23%</td>
<td>4.13%</td>
<td>4.70%</td>
</tr>
<tr>
<td></td>
<td>9.34</td>
<td>4.13</td>
<td>3.11</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>1.86</td>
<td>0.84</td>
<td>1.64</td>
<td>2.53</td>
</tr>
<tr>
<td>6</td>
<td>6.41%</td>
<td>2.31%</td>
<td>3.89%</td>
<td>5.39%</td>
</tr>
<tr>
<td></td>
<td>10.79</td>
<td>4.62</td>
<td>3.57</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>1.89</td>
<td>0.87</td>
<td>1.54</td>
<td>2.68</td>
</tr>
</tbody>
</table>
4.5 Learning from predictability

The empirical section provides robust evidence that monetary policy path shocks predict excess returns on bonds. This section investigates two economic channels through which path shocks may drive the time variation of bond risk premia, and asks which channel is rejected by the data. In general, risk premia are time varying if either the price or quantity of risk are not constant. Indeed the literature has proposed two benchmark economies that give rise to time varying risk premia: (i) habit and (ii) long-run risk (LRR) economies. Can the predictability results found in the previous section be understood within the context of one of these two streams of models?

In Lucas economies with habit preferences as in Campbell & Cochrane (1999), predictability arises in equilibrium because of an endogenously time-varying price of risk. Shocks to the current endowment affect the wedge between consumption and habit, thus inducing time-variation in the price of risk. The impact of consumption shocks on interest rates depends on the relative importance of the intertemporal consumption smoothing and precautionary savings effects. A negative shock to consumption tends to increase the short term interest rate through the consumption smoothing channel: agents expect surplus consumption to recover, so they borrow more against future consumption to smooth their consumption path. At the same time, the shock tends to reduce interest rates through a precautionary savings channel: the conditional volatility of surplus rises when its level drops, inducing agents to save more. Campbell & Cochrane (1999) parametrize the process for surplus in a way that ensures that the two effects offset each other, so that short rates are constant and the term structure of interest rates is always flat. Wachter (2006) removes this restriction and solves for bond prices. She finds that when the model is calibrated to the data, inter-temporal smoothing dominates precautionary savings: the correlation between surplus and the short rate is negative. In this context investors require compensation to hold bonds, which varies depending on the level of consumption surplus.\footnote{Another example of habit economy is Buraschi & Jiltsov (2007), who model a continuous time economy with habit formation and obtain a closed form solution for (non-affine) bond yields.}

In long run risk economies with Epstein & Zin (1989) recursive preferences such as Bansal & Yaron (2004), predictability stems from the dynamics of the quantity of risk: the second
moments for the conditional growth rate of nominal and real macroeconomic variables are assumed to be time varying. The most recent chapter that studies the implications of this class of models in bond markets is Bansal & Shaliastovich (2013). They derive a term structure model for nominal yields based on Epstein & Zin (1989) preferences, time varying expected consumption growth and inflation, time varying volatility of expected consumption growth and inflation, and money non-neutrality. In equilibrium, the impact of uncertainty on bond risk premia depends on the interaction of the sensitivity of yields to expected consumption and inflation growth and the sign of the prices of risk attached to consumption and inflation uncertainty. Nominal bond yields are increasing in both expected consumption and inflation growth. The price of expected consumption risk is positive when agents have preference for early resolution of uncertainty. Since high expected inflation signals low expected consumption growth (via money non-neutrality), the price of expected inflation risk is negative. As a consequence, while high expected growth uncertainty lowers bond premia, high expected inflation uncertainty raises bond premia.

The potential channels through which monetary path shocks can generate bond return predictability are very different in these two classes of models. Consider habit economies first. Habit economies need to feature two ingredients for expected monetary policy shocks to be a driver of risk premia: (i) monetary policy must have an impact on consumption; (ii) habit must be internal. In benchmark habit economies, in fact, the dynamics of consumption are independent of monetary policy shocks. Extending the consumption dynamics from an i.i.d. process (as in Campbell & Cochrane (1999)) to accommodate money non-neutrality, however, is not sufficient: in external habit models, risk aversion is a function of current surplus, which in turn depends on past consumption. Thus, even if monetary policy path shocks affected future consumption, they would not affect current risk aversion and, therefore, the price of risk. In models with internal habit, on the other hand, the current marginal utility of agents also depends on future levels of consumption and habit. Thus, to the extent that expected monetary policy shocks have an impact on future consumption, they can arise as a state variable for the current level of risk aversion and price of risk. The Appendix contains a simple example that shows how money non-neutrality and internal habit can interact to

\footnote{For instance, money non-neutrality may be modeled by assuming that the level of monetary policy shocks enters the conditional mean of consumption growth.}
induce a dependence of risk aversion on monetary policy shocks.

Next, consider a LRR economy. In LRR economies, predictability stems from macroeconomic uncertainty, i.e. heteroskedasticity in the conditional mean of consumption growth and/or inflation. For policy shocks to predict returns via a LRR channel, therefore, it must be the case that monetary policy shocks have an impact on macroeconomic uncertainty. The link between monetary policy and uncertainty has been at the centre of a heated debate: many economists, for instance, have called for the use of policy instruments and communication to explicitly target the extent of macroeconomic uncertainty. This argument has influenced the decision of introducing elements of forward guidance in the toolkit of several central bankers.

To study the implications of these two classes of models in the context of the $PathShocks$-predictability results, we proceed in two steps. First, we compare the empirical evidence in support for a habit versus LRR economy in our sample (independently of the potential existence of monetary channel). Second, we try to understand whether the predictability we observe is consistent with a habit versus LRR channel. To this end we run two sets of regressions. In the first set of regressions, we project consumption surplus onto lagged monetary policy shocks; in the second set of regressions, we regress proxies of macroeconomic uncertainty onto contemporaneous $PathShocks$. Under the null hypothesis that a habit mechanism is at work (time-variation in the price of risk), the slope coefficient in the first regression should be negative and significant. On the other hand, if monetary path shocks generate predictability because of a long-run risk mechanism (time-variation in the quantity of risk) the slope coefficient of the second set of regressions should be positive and significant.

We follow Wachter (2006) and construct a proxy of consumption surplus $s_t$ as a weighted average of 10 years of monthly consumption growth rates:\footnote{The consumption data we use consist of seasonally adjusted, real per-capita consumption of nondurables and services (from the Bureau of Economic Analysis).}

$$s_t = \sum_{j=1}^{120} \phi^j \Delta c_{t-j},$$
where the weight is set to \( \phi = 0.97^{1/3} \) to match the quarterly autocorrelation of the \( P/D \) ratio in the data, as in Wachter (2006); Figure 4.11 plots the dynamics of the series in our sample.

**Figure 4.11: Habit**
This figure plots a proxy of consumption surplus, \( s_t \), defined as:

\[
s_t = \sum_{j=1}^{120} \phi^j \Delta c_{t-j},
\]

where \( \phi = 0.97^{1/3} \) and \( \Delta c_t \) is the (log) consumption growth between months \( t - 1 \) and \( t \). Consumption data consist of seasonally adjusted, real per-capita consumption of nondurables and services.

In order to construct proxies of GDP and inflation uncertainty, we use an ARMA(1, 1)
model to demean consensus expectations about 1-year ahead GDP and CPI growth, and fit a GARCH(1, 1) to the residuals: we use the conditional volatilities implied by the GARCH model as proxies of uncertainty.\textsuperscript{21} Table 4.9 presents the estimates of the ARMA(1, 1)-GARCH(1, 1) model, while Figures 4.12 and 4.13 plot the time series of the conditional volatility of GDP and inflation implied by the estimates.

**Table 4.9: Macroeconomic uncertainty**

This table shows the point estimates and t-statistics of an ARMA(1, 1)-GARCH(1, 1) model for consensus expectations about 1-year-ahead GDP and inflation.

<table>
<thead>
<tr>
<th></th>
<th>ARMA(1, 1)</th>
<th>GARCH(1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>const</td>
<td>AR</td>
</tr>
<tr>
<td>$g^e_{C,t,1Y}$</td>
<td>0,04</td>
<td>0,98</td>
</tr>
<tr>
<td></td>
<td>0,50</td>
<td>35,03</td>
</tr>
<tr>
<td>$\pi^e_{C,t,1Y}$</td>
<td>0,11</td>
<td>0,96</td>
</tr>
<tr>
<td></td>
<td>3,66</td>
<td>89,62</td>
</tr>
</tbody>
</table>

Let us first compare the general predictive properties of these two class of models for bond returns. Table 4.10 shows the results for predictability regressions of bond excess returns on consumption surplus (left panel), macro economic uncertainty (central panel), and both consumption surplus and macro economic uncertainty (right panel). Consider first the predictive ability of habit proxies (left panel). The negative sign of the slope coefficients is intuitive: consistent with the empirical evidence that term premia are countercyclical, an increase in surplus consumption predicts a decline in future excess returns. The statistical significance, however, is weak: the null hypothesis of no significance cannot be rejected for any bond maturity. Next, consider the predictive ability of long run risk proxies (central panel). While inflation uncertainty has no forecasting ability for future excess returns, uncertainty

---

\textsuperscript{21} The construction of the proxies for conditional volatility is closely related to the methodology followed by Bansal & Shaliastovich (2013). Bansal & Shaliastovich (2013) use a VAR to demean consensus expectations about GDP growth and inflation, and then regress the sum of the squared residuals between $t$ and $t + 12$ onto time-$t$ yields; the fitted values are used as a proxies of conditional variances. Our construction allows for the possibility of unspanned macro uncertainty.
Figure 4.12: GDP expectations

This figure plots the consensus 1-year ahead GDP forecast, $g_{CTt+1|Y}$, alongside its conditional volatility, $\sigma_t(g_{CTt+1|Y})$, implied by an ARMA(1, 1)-GARCH(1, 1) model.

about GDP growth features large economic and statistical significance. Contrary to Bansal & Shaliastovich (2013), we find that uncertainty about GDP growth increases, rather than decreases, bond risk premia. The right panel confirms these results by jointly considering both habit and LRR proxies.

Let us now turn to investigate the potential role played by $PathShock$. Table 4.11 reports the output of regressions of consumption surplus on lagged monetary policy path shocks (since the results are robust across specifications of $PathShock$, we only report the estimates for Specification 1). For any horizon considered ($h = 1, \ldots, 5$ years), $PathShock$ fails to predict future levels of consumption surplus, thus ruling out a habit channel. Table
4.12, on the other hand, reports the output of regressions of macroeconomic uncertainty on contemporaneous monetary policy path shocks. The most important result in this table is that path shocks are strongly related to GDP growth uncertainty: the t-statistic rejects the null of no significance at any standard level, and the $R^2$ suggests that as much as one fifth of the variance of RGD uncertainty is accounted for the variation in $PathShock$. These results are consistent with two interpretations: (i) $PathShock$ predicts future excess returns via a quantity of risk channel; (ii) more precisely, $PathShock$ impacts the uncertainty surrounding GDP growth, which, consistent with a LRR economy, is a key driver of the time variation of bond risk premia in our sample (see Table 4.10). While the link between $PathShock$
Table 4.10: Excess returns, habit and LRR

The table reports the output from regressions of annual bond excess returns on proxies \( x_t \) of consumption surplus and long run risk: \( r_{t+12}^{(n)} = \text{const.} + \beta^{(n)} x_t + \epsilon_{t+12}^{(n)} \). Bond maturities \( n \) range from 2 to 5 years. The proxies are: consumption surplus \( (s) \), GDP growth uncertainty \( (\phi) \), inflation uncertainty \( (\pi^e) \). T-statistics, reported below the point estimates, are corrected for auto-correlation and heteroskedasticity using Newey-West errors (18 lags). \( \bar{R}^2 \) is the adjusted \( R^2 \). Both left and right hand variables are standardized. A constant is included but not reported.

<table>
<thead>
<tr>
<th>Habit</th>
<th>LRR</th>
<th>Habit and LRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>( s )</td>
<td>( \bar{R}^2 )</td>
</tr>
<tr>
<td>2 -0.08</td>
<td>0.22%</td>
<td>0.37</td>
</tr>
<tr>
<td>-0.46</td>
<td></td>
<td>3.03</td>
</tr>
<tr>
<td>3 -0.09</td>
<td>0.39%</td>
<td>0.41</td>
</tr>
<tr>
<td>-0.50</td>
<td></td>
<td>3.64</td>
</tr>
<tr>
<td>4 -0.14</td>
<td>1.47%</td>
<td>0.42</td>
</tr>
<tr>
<td>-0.87</td>
<td></td>
<td>3.82</td>
</tr>
<tr>
<td>5 -0.17</td>
<td>2.46%</td>
<td>0.41</td>
</tr>
<tr>
<td>-1.34</td>
<td></td>
<td>3.98</td>
</tr>
</tbody>
</table>

and \( \sigma^e (\phi) \) is tight, there is a large fraction of variation in \( \text{PathShock} \) that is unrelated to fluctuations in RGD uncertainty. It is natural to ask what is the predictive power of the component of \( \text{PathShock} \) that is unrelated to GDP uncertainty for bond excess returns. We address this question by running predictive regressions on \( \text{PathShock} \) in which we control for macroeconomic uncertainty. Table 4.13 summarizes the results. While controlling for uncertainty does indeed reduce the forecasting power of \( \text{PathShock} \), path shocks retain significant forecasting power at all bond maturities.
Table 4.11: Monetary policy shocks and surplus

The table reports the output from regressions of consumption surplus at time $t + h$ ($s_{t+h}$) on expected monetary policy shocks (specification 1) at time $t$: $s_{t+h} = const. + \beta PathShock_t + \epsilon_{t+h}$. Forecasting horizons ($h$) range from 1 to 5 years. $T$-statistics, reported below the point estimates, are corrected for auto-correlation and heteroskedasticity using Newey-West errors (lags equal to $h$). $\bar{R}^2$ is the adjusted $R^2$. Both left and right hand variables are standardized. A constant is included but not reported.

<table>
<thead>
<tr>
<th>$h$</th>
<th>PathShock</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.07</td>
<td>0.10%</td>
</tr>
<tr>
<td></td>
<td>-0.44</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.02</td>
<td>-0.36%</td>
</tr>
<tr>
<td></td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.79%</td>
</tr>
<tr>
<td></td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>5.66%</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.13</td>
<td>1.16%</td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.12: Monetary policy shocks and uncertainty

The table reports the output from regressions of macroeconomic (GDP, inflation) uncertainty at time $t$ ($\sigma_t(\cdot)$) on monetary policy path shocks (specification 1) at time $t$: $\sigma_t(\cdot) = \text{const.} + \beta \text{PathShock}_t + \epsilon_t$. T-statistics, reported below the point estimates, are corrected for autocorrelation and heteroskedasticity using Newey-West errors (18 lags equal). $\bar{R}^2$ is the adjusted $R^2$. Both left and right hand variables are standardized. A constant is included but not reported.

<table>
<thead>
<tr>
<th>PathShock</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma (\sigma^e)$</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>3.27</td>
</tr>
<tr>
<td>$\sigma (\pi^e)$</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
</tr>
</tbody>
</table>
4.6 Conclusion

Motivated by the monetary policy underpinnings of Treasury yields, this chapter asks whether monetary policy shocks represent an important source of time variation in bond risk premia; more specifically, it accomplishes three tasks. First, it constructs PathShock, a measure of exogenous variation in the path of monetary policy, from the residuals of Taylor rules estimated on survey forecasts of federal funds rates, GDP growth, and inflation. Second, it investigates whether PathShock captures the time variation in bond excess returns. Third, it studies whether the results are consistent with time variation in the quantity of risk, or price of risk.

Indeed, we find that monetary policy shocks account for 10%-15% of the variance of one-year excess returns for bond maturities between 2 and 5 years. The predictability is economically and statistically strong even after controlling for the levels of macro economic activity; the results are robust to the choice of Taylor rule specification, and to the inclusion of the financial crisis in the sample. Furthermore, we find that the evidence on predictability is consistent with a quantity of risk channel: in particular, PathShock increase the conditional volatility of GDP growth, a source of time varying risk premia in long run risk economies.

The results have implications for both asset pricing and monetary policy. First, the evidence suggests that macro-finance models seeking to understand the role of monetary policy in financial markets should introduce money non-neutrality in the form of a dependence of macroeconomic uncertainty onto monetary policy shocks. Second, the results indicate that exogenous variation in forward guidance impacts bond prices via risk premia: policy makers should be aware of this transmission channel when implementing policy.

4.7 Appendix

4.7.1 Estimation of path shocks

This section discusses different approaches to estimate monetary policy path shocks. For ease of notation, re-write the model for panel data as:

$$y_{nt} = x_{nt}'\beta + \alpha_n + \epsilon_{nt}.$$
where

\[ y_{nt} = f_{n,t,h}^e \]
\[ x'_{nt} = \begin{bmatrix} 1 & f_{n,t,h-1}^e & \cdots & f_{n,t,h-k}^e & \pi_{n,t,h+j}^e & x_{n,t,h+k}^e \end{bmatrix}' \]
\[ \epsilon_{nt} = u_{n,t,h}^e. \]

\( \alpha_n \) is an agent-specific term that controls for possible cross-sectional unobservable heterogeneity. A structural interpretation of \( \alpha_n \) may be, for instance, that agents have different opinions about the long run mean of the federal funds rate, or, alternatively, about the inflation target of the Central Bank. The assumption that monetary policy shocks are orthogonal to the arguments of the Taylor rule implies:

\[ E[\epsilon_{nt}|x_{nt}] = 0. \]

Unobserved heterogeneity, on the other hand, may be correlated (mean dependence) or not correlated (mean independence) with the arguments of the Taylor rule.

The response coefficient \( \beta \) can be estimated from forecast data in at least four ways: (i) using consensus data, via OLS; or, using panel data, via: (ii) pooled OLS (POLS); (iii) fixed effects (FE); (iv) random effects (RE). It is natural to ask what are the relative advantages and disadvantages of each estimator, and under which conditions they provide the same answer.

Consider first the case of mean independence. Stack the \( N \) cross-sectional observations at time \( t \) into a single equation:

\[ y_t = x_t' \beta + \alpha + \epsilon_t; \]

the POLS estimator is given by:

\[ \hat{\beta}_P = \left( \sum_{t=1}^T x_t' x_t \right)^{-1} \sum_{t=1}^T x_t' y_t \]
\[ = \left( \sum_{t=1}^T x_t' x_t \right)^{-1} \sum_{t=1}^T x_t'(x_t \beta + \alpha + \epsilon_t) \]
\[ = \beta + \left( \sum_{t=1}^T x_t' x_t \right)^{-1} x_t' \alpha + \left( \sum_{t=1}^T x_t' x_t \right)^{-1} x_t' \epsilon_t. \]
Notice that the term
\[ E \left[ \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} x'_t \alpha \right] \]
is zero under the assumption of mean independence, and non-zero otherwise. Also, under
the assumption \( E [\epsilon_{nt} | x_{nt}] = 0 \), an application of the law of iterated expectations implies:
\[
E \left[ \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} x'_t \epsilon_t \right] = E \left[ E \left[ \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} x'_t \epsilon_t | x_t \right] \right]
= E \left[ \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} x'_t E [\epsilon_t | x_t] \right]
= 0.

The implications is that the POLS estimator is unbiased only if mean independence holds;
if, on the other hand, heterogeneity is correlated with the regressors, the POLS estimator is
biased and a FE estimator should be used.

Now let variables without \( n \)-subscripts and with over-bars denote cross-sectional averages,
and consider the consensus model:
\[
\bar{y}_t = \bar{x}_t \beta + \bar{\alpha} + \bar{\epsilon}_t;
\]
By following the same steps outlined above, it can be shown that the consensus estimator
can be written as:
\[
\hat{\beta}_C = \beta + \left( \sum_{t=1}^{T} \bar{x}'_t \bar{x}_t \right)^{-1} \bar{x}'_t \bar{\alpha} + \left( \sum_{t=1}^{T} \bar{x}'_t \bar{x}_t \right)^{-1} \bar{x}'_t \bar{\epsilon}_t.
\]
Once more, the assumption of mean independence implies:
\[
E \left[ \left( \sum_{t=1}^{T} \bar{x}'_t \bar{x}_t \right)^{-1} \bar{x}'_t \bar{\alpha} \right] = 0;
\]
124
however, we have that:

\[
E \left[ \left( \sum_{t=1}^{T} \bar{x}_t \bar{x}_t \right)^{-1} \bar{x}_t \bar{\varepsilon}_t \right] = E \left[ E \left[ \left( \sum_{t=1}^{T} \bar{x}_t \bar{x}_t \right)^{-1} \bar{x}_t \bar{\varepsilon}_t | \bar{x}_t \right] \right] \\
= E \left[ \left( \sum_{t=1}^{T} \bar{x}_t \bar{x}_t \right)^{-1} \bar{x}_t E \left[ \bar{\varepsilon}_t | \bar{x}_t \right] \right] \\
\neq 0,
\]

implying that, even under the assumptions of mean independence, the consensus estimator is biased unless

\[
E \left[ \bar{\varepsilon}_t | \bar{x}_t \right] = E \left[ \varepsilon_{nt} | x_{nt} \right] = 0.
\]

Should this condition hold, the consensus estimator would be unbiased, albeit inefficient compared to the pooled OLS estimator.

In summary, the choice of the estimator depends on the assumptions about the orthogonality between cross-sectional average of the forecasts about the arguments of the Taylor rule and the cross-sectional average of monetary shocks forecasts, on one hand, and unobserved heterogeneity, on the other hand. Absent economic priors, it is possible to dispense with the need to make an assumption about either condition, and rely on a purely statistical criteria for the selection of the estimator. In particular, the null hypothesis of no fixed effects can be tested by running an F-test on the hypothesis of joint significance of the agent dummies. The null hypothesis of random effects, on the other hand, can be tested by means of a Hausman test.

### 4.7.2 Money non-neutrality and habit: an example

This section presents a simple example of how money non-neutrality and internal habit can interact to generate a link between monetary policy shocks and risk aversion. The intuition is simple. Internal habit implies that agents are concerned about the level of future consumption surplus which, under money non-neutrality, is sensitive to monetary policy shocks: since the local curvature of utility is inversely related to consumption surplus, monetary policy
shocks have an impact on the coefficient of relative risk aversion and, therefore, on the price demanded by agents to bear a unit of risk.

A simple way to model money non-neutrality in a reduced form way is to allow lagged monetary policy shocks to affect the conditional mean of consumption growth. Let $C_t$ and $u_t$ denote the level of aggregate consumption and monetary policy shocks, respectively. Assume that the dynamics of $\Delta c_{t+1} = \log \left( \frac{C_{t+1}}{C_t} \right)$ and $u_t$ are described by:

$$\Delta c_{t+1} = -\kappa u_t + \sigma e^u_t$$
$$u_{t+1} = \phi_u u_t + \sigma u^u_t.$$  

The parameter $\kappa$ controls the extent of money non-neutrality. If $\kappa = 0$ money neutrality holds: consumption growth is iid with mean zero and monetary policy has no effect on consumption. On the other hand, if $\kappa \neq 0$, money is non-neutral. Suppose, for instance, that the monetary cycle is contractionary ($u_t > 0$): the higher $\kappa$, the lower is the conditional mean of consumption growth. The AR(1) assumption for $u_t$ reflects the mean-reverting nature of monetary policy shocks, and implies that the effect of monetary policy on the conditional growth of consumption is transitory: monetary policy cannot permanently affect the growth rate of the economy. Clearly, this process for consumption is only a simplistic representation; the process can be straightforwardly extended to include, for instance, a non-zero unconditional mean and additional non-monetary sources of variation in the conditional mean.

The representative agent maximizes

$$U_t (C_t, C_{t+1}, \ldots) = E_t \left[ \sum_{j=0}^{\infty} e^{-\delta j} u_{t+j} (C_{t+j}, H_{t+j}) \right],$$

where $u_t (C_t, H_t)$ is a function that maps time-$t$ consumption ($C_t$) and habit ($H_t$) into time-$t$ utility. A common assumption in the literature (see, for instance, Campbell & Cochrane (1999), Wachter (2006), Buraschi & Jiltsov (2007), Santos & Veronesi (2010)) is that $u_t (C_t, H_t)$ is a CRRA utility function defined over the difference between consumption and habit:

$$u_{t+j} (C_{t+j}, H_{t+j}) = \frac{\left( C_{t+j} - H_{t+j} \right)^{1-\gamma}}{1-\gamma}.$$
In general, habit is a function of current and past consumption:

\[ H_t = f(C_t, C_{t-1}, C_{t-2}, \ldots) \]

typically, it is defined as an exponentially weighted average.

The main object of interest of this section is to understand how the coefficient of relative risk aversion

\[ \eta_t = -\frac{C_t}{U_c} \frac{U_{cc}}{U_c}, \]

varies as a function of monetary policy shocks. In general, the first and second derivative of utility with respect to consumption are given by:

\[
U_c = \frac{\partial u_t(C_t, H_t)}{\partial C_t} + E_t \left[ \sum_{j=1}^{\infty} e^{-\delta j} \frac{\partial u_{t+j}(C_{t+j}, H_{t+j}) \partial H_{t+j}}{\partial H_{t+j}} \right],
\]

\[
U_{cc} = \frac{\partial^2 u_t(C_t, H_t)}{\partial C_t^2} + E_t \left[ \sum_{j=1}^{\infty} e^{-\delta j} \frac{\partial^2 u_{t+j}(C_{t+j}, H_{t+j}) \partial H_{t+j}}{\partial C_t} \right].
\]

With external habit, agents ignore the impact of today’s consumption on future habits \((\partial H_{t+j}/\partial C_t = 0)\), so the expectation terms are zero and the first and second partial derivatives simplify to \(\partial u_t(C_t, H_t)/\partial C_t\) and \(\partial^2 u_t(C_t, H_t)/\partial C_t^2\), respectively. Since monetary policy affects the real economy (consumption) only with a lag \((u_t\) affects \(C_{t+j}\) but not \(C_t\)), it is clear that RRA is independent of monetary policy shocks when when habit is external. Under internal habit, on the other hand, agents do not ignore the impact that current consumption has on future habit \((\partial H_{t+j}/\partial C_t \neq 0)\), so that \(U_c\) and \(U_{cc}\) retain the general form illustrated above and RRA depends on \(u_t\) via \(C_{t+j}\).

In order to illustrate the relationship between \(u_t\) and \(\eta_t\), assume that utility is defined over the ratio of consumption to habit and that habit is a (linear) function of lagged consumption:

\[
U_t(C_t, C_{t+1}, \ldots) = E_t \left[ \sum_{j=0}^{\infty} e^{-\delta j} \frac{(C_{t+j}/H_{t+j})^{1-\gamma}}{1 - \gamma} \right]
\]

\[ H_t = hC_{t-1}. \]
Modeling the wedge between consumption and habit as a ratio implies a constant coefficient of risk aversion under external habit but not under internal habit. We choose the specification where utility is a function of $C_t/H_t$, instead of $C_t - H_t$, because of analytical tractability (we can compute the expectation terms in the former case). The specification for the habit function is highly unrealistic, but it serves the purpose to simplify calculations: $\partial^m H_{t+j}/\partial C_t^n \neq 0$ only for $j = 1$ and $n = 1$, so the impact of current consumption on future habits is limited to first order effects and to the first horizon. Under these assumptions the expression for RRA can be computed in closed form. The first and second derivative of utility are:

$$U'_C = (C_t/H_t)^{-\gamma} H_t^{-1} - E_t \left[ \sum_{j=0}^{\infty} e^{-\delta_j} C_t^{1-\gamma} H_t^{-(2-\gamma)} \frac{\partial H_{t+j}}{\partial C_t} \right]$$

$$= (C_t/H_t)^{-\gamma} H_t^{-1} - E_t \left[ e^{-\delta} C_t^{1-\gamma} H_t^{-(2-\gamma)} \right]$$

$$= (C_t/H_t)^{-\gamma} H_t^{-1} - E_t \left[ e^{-\delta} (hC_t)^{-(2-\gamma)} h \right]$$

$$= (C_t/H_t)^{-\gamma} H_t^{-1} - e^{-\delta} (hC_t)^{-(2-\gamma)} h E_t \left[ C_t^{1-\gamma} \right]$$

$$= (C_t/H_t)^{-\gamma} H_t^{-1} - e^{-\delta} (hC_t)^{-(2-\gamma)} h C_t^{(1-\gamma)} e^{-(1-\gamma)\kappa\mu_t + 0.5(1-\gamma)^2\sigma^2}$$

$$= (C_t/H_t)^{-\gamma} H_t^{-1} - e^{-\delta} \gamma C_t^{-1} e^{-(1-\gamma)\kappa\mu_t + 0.5(1-\gamma)^2\sigma^2}$$

$$U''_{CC} = -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2}$$

$$- E_t \left[ \sum_{j=0}^{\infty} e^{-\delta_j} C_t^{1-\gamma} \left( - (2-\gamma) H_t^{-(3-\gamma)} \left( \frac{\partial H_{t+j}}{\partial C_t} \right)^2 + H_t^{-(2-\gamma)} \frac{\partial^2 H_{t+j}}{\partial C_t^2} \right) \right]$$

$$= -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2} - E_t \left[ e^{-\delta} C_t^{1-\gamma} \left( - (2-\gamma) H_t^{-(3-\gamma)} \left( \frac{\partial H_{t+1}}{\partial C_t} \right)^2 \right) \right]$$

$$= -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2} - E_t \left[ e^{-\delta} C_t^{1-\gamma} \left( - (2-\gamma) hC_t^{-(3-\gamma)} h^2 \right) \right]$$

$$= -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2} + (2-\gamma)e^{-\delta} (hC_t)^{-(3-\gamma)} h^2 E_t \left[ C_t^{1-\gamma} \right]$$

$$= -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2} + (2-\gamma)e^{-\delta} (hC_t)^{-(3-\gamma)} h^2 C_t^{(1-\gamma)} e^{-(1-\gamma)\kappa\mu_t + 0.5(1-\gamma)^2\sigma^2}$$

$$= -\gamma (C_t/H_t)^{-\gamma-1} H_t^{-2} + (2-\gamma)e^{-\delta} h^{\gamma-1} C_t^{-2} e^{-(1-\gamma)\kappa\mu_t + 0.5(1-\gamma)^2\sigma^2}$$

22The former result partly explains the assumption, common in external habit models, that utility is a function of the difference between consumption and habit.
where we have used the fact:

\[ E_t [C_{t+1}] = C_t^\beta e^{-\delta\kappa u_t + 0.5\delta^2\sigma^2}, \]

which follows directly from the dynamics for \( C_t \) and the assumption of lognormality. It follows that the RRA is equal to:

\[ \eta_t = C_t \gamma \left( \frac{C_t}{H_t} \right)^{-\gamma-1} H_t^{-2} + (\gamma - 2) e^{-\delta h\gamma-1} C_t^{1-\gamma} e^{-\gamma-1} C_t e^{-0.5(1-\gamma)^2\sigma^2} \]

\[ \frac{1}{C_t H_t^{-\gamma - 1} - e^{-\delta h\gamma-1} C_t^{1-\gamma} e^{-\gamma-1} C_t e^{-0.5(1-\gamma)^2\sigma^2}} \]

Figure ?? shows the coefficient of RRA as a function of monetary policy shocks. The plot is based on the following assumptions: (i) CRRA curvature: \( \gamma = 5 \); (ii) time preference: \( \delta = 0 \); (iii) non-neutrality parameter: \( \kappa = 4 \) (a 25 basis points monetary policy shock reduces \( E_t[\Delta C_{t+1}] \) by 1%); (iv) initial consumption and habit: \( C_t = H_t = 1 \) (all the variation in RRA stems from expectation terms); (v) habit parameter \( h = 0.75 \) (low enough to ensure that consumption does not fall below habit); (vi) consumption volatility: \( \sigma_c = 1\% \). The magnitude of monetary policy shocks encompasses the range of potential rate cuts/increases: \( u_t = 0, \pm 25bp, \pm 50bp, \pm 75bp \). The figure shows that monetary policy shocks have a non-trivial effect on relative risk aversion: the price of risk is an increasing function of \( u_t \).
Figure 4.14: Monetary policy shocks and RRA
This figure plots the coefficient of relative risk aversion as a function of monetary policy shocks for an economy with internal habit and money non neutrality.
Chapter 5

Conclusion

This thesis studies return predictability in both equity (chapters 2 and 3) and fixed income markets (chapter 4). Chapter 2 shows that time varying risk premia cannot be easily exploited in real time, even though evidence of in-sample predictability is strong. We reach this conclusion by comparing the out-of-sample performance of two portfolios: a portfolio that shifts its equity exposure based on dividend yield signals, and one that does not. In general, the dynamics of optimal portfolio weights are a function of the time variation in both conditional expected returns and hedging demand. By assuming that the horizon of the fund manager is annual, we do not take into account the variation in weights due to changes in the hedging demand component. An interesting extension would be to value the economic gains accruing to investors with longer time horizons. To the extent that the significance of long run predictability varies over time, longer horizons may have non trivial effects on tactical portfolio allocation via the hedging demand channel.

Chapter 3 explores the dynamics of equity risk premia on equity claims of different maturity. First, we construct the prices of synthetic dividend claims with a maturity of one year. Next, we obtain a measure of short maturity risk premia by projecting the returns of dividend claims on their lagged price-to-dividend ratios. Finally, we explore the link between short and long maturity risk premia by assessing the predictive power of short maturity premia for the returns on the index. We find that high risk premia on short maturity claims predict index returns with a negative sign. This result is puzzling, because it cannot be reconciled with the predictions of leading asset pricing models such as long run risk and habit economies. Our
study contributes to a fast growing research area that seeks to understand the economics of dividend claims. We believe that this research area should have high priority in the asset pricing agenda for two reasons. First, the recent growth of dividend markets calls for a better understanding of the economics and pricing of short maturity equity claims. Second, observations on dividend prices provide additional empirical restrictions useful to the testing of asset pricing models. The empirical results uncovered by chapter 3 are puzzling, suggesting that a lot of work still needs to be done on the theoretical front. A possible direction for future research is to explore a long run risk model with multiple volatility factors that operate at different frequencies.

Chapter 4 investigates the interaction of monetary policy and bond risk premia. We use a novel identification scheme to construct a measure of monetary policy shocks from survey data, and document its considerable forecasting power for holding period excess returns on treasury securities. We ask whether the empirical evidence is consistent with a quantity of risk or price of risk channel, and find that monetary policy path shocks increase the conditional volatility of GDP growth, a source of time varying risk premia in long run risk economies. The findings have important implications for the transmission mechanism of monetary policy: contractionary policies, besides slowing down output growth and inflation, may also raise GDP uncertainty and risk premia. We believe that a promising avenue for future research is to explore the full equilibrium implications of a long run risk economy with money non neutrality.

Time variation in expected returns is only the observable manifestation of an invisible realm of complex preferences and rich risk dynamics. Our humble hope is that this thesis has been successful at contributing to the understanding of this fascinating phenomenon.
Bibliography


