The Role of Fundamentals in Asset Pricing

Ian Roberts
Imperial College Business School
Ph.D Finance Thesis

This version: 3rd June 2014
Abstract

This thesis seeks to contribute to the understanding and measurement of the aggregate macroeconomic risks that drive security prices, via three empirical studies. Following Abreu and Brunnermeier (2002), Chapter 2 examines a wide cross-section of global equity markets to test the hypothesis that not only do deviations from fundamentals occur, but it is optimal for an arbitrageur to delay trades that would correct the mispricing. The evidence supports the hypothesis that market prices can and do experience sustained deviations from fundamentals despite the presence of arbitrageurs. Chapter 3 examines the tangled evidence that relates firm-level financial distress to the market, size and value factors of Fama and French (1996). Using panel data it is then shown that exposure to the Fama-French factors can be linked to distress risk, which demonstrates a link between priced aggregate macroeconomic risks and the financial performance of individual firms. Frazzini & Pedersen (2013)’s Betting-Against-Beta (BAB) factor is theorized to occur due to the leverage constraints of a subset of the population of investors. Chapter 4 demonstrates that in the post-1990 subsample the premium on the BAB factor is significantly diminished. Furthermore a conditional model using Carhart (1997)’s 4-factors reduces the alpha further, suggesting that part of the anomaly is in fact attributable to conditional exposure to existing factors, rather than market frictions.
Declaration

This work is my own and consists of original content except where indicated by references. It has been submitted for consideration for the degree of Ph.D. in Finance at Imperial College Business School, and was completed under the guidance of Professor Walter Distaso and Professor James Sefton.

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.
Acknowledgements

I would like to take the opportunity to thank my supervisors Professor Walter Distaso and Professor James Sefton, who have provided valuable academic support throughout this journey. Professor Marcelo Fernandes, Professor Giulia Iori, Dr. Timothy French and Dr. Benjamin Voyer all generously offered me their time and encouragement during the embryonic stages of this thesis. Conversations with Christian Deazeley and Graeme Blyth have been an indirect inspiration during my progress as a researcher, and I have always valued their insights as financial practitioners.

Finally I would like to extend a special thank you to my parents, Raymond and Jeyalakshmi Roberts, both of whom have always provided me with their loving support.
“Life shrinks or expands in proportion to one’s courage.”

- Anais Nin.
Contents

Chapter 1: Introduction . . . . . . 10

Chapter 2: Evidence for Limited Arbitrage in Global Equity Markets . . . . . . 15

Chapter 3: Distress Risk in the Size and Value Factors . . . . . . 55

Chapter 4: Low Risk Investing . . . . . . 81

Chapter 5: Conclusions of Thesis . . . . . . 104

Bibliography . . . . . . 106

Figures . . . . . . 114
List of Tables

1. Trading Strategy Performance Metrics for V1. ........................................ 31
2. Stepwise Multiple Hypothesis Tests for V1 Using Log Utility. .................. 32
3. Stepwise Multiple Hypothesis Tests for V1 Using Sharpe Ratios. ................. 33
4. Stepwise Multiple Hypothesis Tests for V2 Using Log Utility. .................. 37
5. Stepwise Multiple Hypothesis Tests for V2 Using Sharpe Ratios. ................. 38
7. Stepwise Multiple Hypothesis Tests for V3 Using Sharpe Ratios. ................. 40
8. Differences-in-Mean T-tests for the Synchronization Risk Preference Parameter. 46
9. Descriptive Statistics for Bankruptcy Dataset. ........................................ 63
10. Descriptive Statistics for the Explanatory Variables. .............................. 65
11. Naive Logit Regressions for Default. .................................................. 67
12. Naive Logit Regressions for Default at Longer Horizons. ......................... 70
13. Default Prediction Using the 4-factor Betas and $SIGMA(R)$. .................... 71
14. SIMEX Logit Regression for Default. .................................................. 74
15. Logit Regressions for Default Using Weekly Return Observations: Non-Synchronous Data. ................................................................. 75
16. Number and Assets of Hedge Funds. .................................................. 88
17. Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta. ................................................................. 90
18. Chow Test for a Structural Break between the 1926-1990 and 1990-2012 Subsamples. ................................................................. 91
19. Decomposition of the Unconditional Alpha. ......................................... 92
20. Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta: Non-Synchronous Data .............................................. 95
21. Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta: Quintile Portfolios .......................................... 96
22. Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta (1985) ................................................................. 102
23 Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta (1994)
# List of Figures

1. V1 OOS Mispricing Episodes for Chile .................................................. 114
2. V1 OOS Mispricing Episodes for Denmark ............................................. 115
3. V1 OOS Mispricing Episodes for Spain .................................................. 116
4. V1 OOS Mispricing Episodes for Finland .............................................. 117
5. V1 OOS Mispricing Episodes for Hong Kong ......................................... 118
6. V1 OOS Mispricing Episodes for Italy .................................................... 119
7. V1 OOS Mispricing Episodes for Philippines ......................................... 120
8. V1 OOS Mispricing Episodes for Sweden .............................................. 121
9. V1 OOS Mispricing Episodes for Thailand ............................................ 122
10. Histogram of Price Refresh Rates ......................................................... 123
11. Histogram of HML Betas for D1 ............................................................ 124
12. Histogram of HML Betas for D2 ............................................................ 125
13. Histogram of Market Betas for D1 ........................................................ 126
14. Histogram of Market Betas for D2 ........................................................ 127
15. Histogram of SMB Betas for D1 ............................................................ 128
16. Histogram of SMB Betas for D2 ............................................................ 129
17. SVM Probability of Default for E1D1 as a Function of HML Beta ............ 130
18. SVM Probability of Default for E1D1 as a Function of MKT Beta ......... 131
19. SVM Probability of Default for E1D1 as a Function of SMB Beta .......... 132
20. SVM Probability of Default for E1D2 as a Function of HML Beta ............ 133
21. SVM Probability of Default for E1D2 as a Function of MKT Beta .......... 134
22. SVM Probability of Default for E1D2 as a Function of SMB Beta .......... 135
23. SVM Probability of Default for E1D2 as a Function of Weekly HML Beta .. 136
24. SVM Probability of Default for E1D2 as a Function of Weekly MKT Beta .. 137
25. SVM Probability of Default for E1D2 as a Function of Weekly SMB Beta .. 138
26. Number of Firms in Sample for V1 ...................................................... 139
Chapter 1
Introduction

"The CAPM and its successor factor models are paradigms of the absolute approach. Yet in applications, they price assets "relative" to the market or other risk factors, without answering what determines the market or factor risk premia and betas... The central and unfinished task of absolute asset pricing is to understand and measure the sources of aggregate or macroeconomic risk that drive asset prices. Of course, this is also the central question of macroeconomics, and this is a particularly exciting time for researchers who want to answer these fundamental questions in macroeconomics and finance. For example, expected returns vary across time and across assets in ways that are linked to macroeconomic variables, or variables that also forecast macroeconomic events; a wide class of models suggests that a "recession" or "financial distress" factor lies behind many asset prices." - Cochrane (2005).

The relationship between fundamental macroeconomic risks and security prices is of paramount importance to finance. However in practice both the Capital Asset Pricing Model\(^1\) and Black-Scholes equation\(^2\) price assets relative to other assets, leaving aside the role of economic fundamentals. The Fama and French (1996) and Carhart (1997) models advance on the CAPM by using portfolios of stocks sorted on firm characteristics as additional factors, but this raises even more questions about the relationship between economic fundamentals and security prices, as it becomes necessary to explain what aggregate macroeconomic risks are proxied by the sorted portfolios. Accordingly there have been many attempts to relate security prices to economic fundamentals, and this thesis aims to contribute to that stream of literature.

Chapter 2 investigates the evidence for limited arbitrage in global equity markets. In the classical theory of finance, arbitrage has the effect of equalizing market prices to fundamental values thus keeping markets efficient. Moreover Friedman (1953) and Fama (1965) suggest

---

\(^1\)Sharpe (1964) and Lintner (1965).
\(^2\)Black and Scholes (1973).
that even in the presence of irrational investors, the actions of rational arbitrageurs will drive prices close to fundamental values. However this argument assumes that arbitrageurs are well-resourced and unconstrained. There is a stream of literature that relaxes these assumptions and considers how arbitrage might be limited in practice. If arbitrage is limited, this has major consequences for any attempt to theoretically link economic fundamentals to market prices. It would mean that even if there were a model used by rational arbitrageurs with homogenous beliefs to price securities as a function of economic fundamentals, it would remain possible for market prices to deviate from these fundamental prices in a sustained way.

In markets that are incomplete arbitrageurs are never able to fully hedge the risk in a position involving a mispriced asset, which prevents them from taking unbounded positions to correct mispricings. This has two consequences, namely fundamental risk and noise trader risk. The former is where the fundamental value of a partially hedged portfolio may change over time, or arbitrageurs may find that their model doesn’t coincide with the true data-generating process. This causes the arbitrageur to suffer losses even if they are able to sustain the strategy until the final payoff is realized [Wurgler and Zhuravskay (2002)]. Shiller (1984) and Campbell and Kyle (1987) show that fundamental risk can severely limit arbitrage even when arbitrageurs have infinite horizons. Noise trader risk is where behavioural traders push prices away from fundamentals, and constrained arbitrageurs cannot correct the mispricing immediately. This may cause arbitrageurs to suffer losses if they are forced to liquidate early [DeLong et al (1990a)]. Consequently if markets are not complete, the presence of some rational arbitrageurs is not enough to correct the mispricings arising from behavioural traders, contrary to the efficient markets hypothesis.

Shleifer and Vishny (1997) note that professional arbitrage is often conducted by specialists using other people’s capital. These rational arbitrageurs might find it optimal to delay taking corrective bets because of the risk that the mispricing worsens during their performance evaluation horizon, leading to underperformance and fund outflows due to the agency problem. Such an outcome could also leave them without capital precisely when the arbitrage opportunities are best. Abreu and Brunnermeier (2002) assume that each individual arbitrageur has finite resources, and therefore more than one arbitrageur needs to make a corrective trade before
any mispricing is removed. In this setting, if any individual arbitrageur is uncertain about when their peers will make their corrective trades, this can incentivize the individual to delay, and since in equilibrium everyone delays, the mispricing persists. I investigate the evidence for limited arbitrage in global equity markets in the context of Abreu and Brunnermeier (2002)’s model. I find evidence that supports the hypothesis that market prices can and do experience sustained deviations from fundamentals despite the presence of arbitrageurs. This is highly consequential to the task of understanding and measuring how economic fundamentals affect security prices. If market valuations are prone to diverge from fundamental values in a sustained way, this makes it more challenging to identify the relationship between security prices and fundamentals.

Chapter 3 investigates whether a stock’s factor loadings from the Fama and French (1996) model are related to firm-level financial distress. Fama and French originally proposed their multifactor model as a way to enhance the CAPM, following the identification of a number of anomalies. These anomalies included the size effect of Banz (1981), the leverage effect of Bhandari (1988), the book-to-market effect of Stattman (1980) and Rosenberg, Reid and Lanstein (1985), and the earnings-price effect of Basu (1983). By including long-short portfolios of stocks sorted on characteristics, SMB and HML, Fama and French’s enhanced multifactor model offered to solve all of these anomalies. Empirically the model worked well in pricing portfolios of stocks sorted according to characteristics. However the success of the model made it important to identify what aggregate macroeconomic risks the SMB and HML factors proxied for. In the CAPM the market return serves as a proxy for returns to the wealth portfolio, which is of hedging concern to investors. Consequently economic theory tells us that the more a security covaries with the market return, the higher the risk premium required by rational risk averse investors. Since the inclusion of SMB and HML in a factor model prices a wider range of portfolios, they must proxy for exposure to aggregate risk factors of hedging concern to investors as well.


Zhang (2005) also links value portfolios to the state of the economy. These studies would seem to corroborate Fama and French’s suggestion that their model is consistent with a multifactor version of Merton (1973)’s ICAPM, in which size and book-to-market proxy for sensitivity to risk factors in returns. These could include unmeasured components of wealth not captured by the market portfolio such as human capital or debt securities, or changes in the investment opportunity set. In the ICAPM, the investment opportunity set becomes relevant to investors because the model advances from the single-period mean-variance optimization specified by Markowitz (1959) to a multiperiod setting. Since the size and book-to-market sorted portfolios SMB and HML are themselves sensitive to innovations in the investment opportunity set, or unmeasured components of wealth, they consequently succeed in pricing other securities that are sensitive to those risk factors.

However as Cochrane (2005) suggests in the quote that started this chapter, there could be a financial distress factor behind many asset prices. In fact it was originally Fama and French (1995) who noted that high book-to-market stocks could be in financial distress due to their sustained low profitability. Dichev (1998) finds that bankruptcy risk (proxied by Z-score and O-score) is negatively related to firm size and positively related to book-to-market. Similarly Campbell et al (2008) find that financially distressed firms have high loadings on the HML and SMB factors. I use a panel data of stocks to find that high magnitudes of loadings on the HML and SMB factors are linked to actual incidence of firm-level default. This result links firm-level financial performance to the aggregate risks that Petkova (2006), Lettau and Ludvigson (2001), Vassalou (2003) and Liew and Vassalou (2000) argue lie behind the Fama and French (1996) factors. Stock returns are sensitive to these aggregate macroeconomic risks when their fundamental financial performance is poor and they are at risk of bankruptcy.

Chapter 4 returns to the CAPM, which of course predicts that the expected excess return on a stock is linearly related to its beta on the market excess return. Frazzini and Pedersen (2013) show that for a sample of US equities 1926-2012, the risk-adjusted returns on portfolios of stocks sorted according to their CAPM beta are monotonically decreasing as the portfolio beta increases. Moreover Baker, Bradley and Wurgler (2010, 2013) show that the geometric mean return on portfolios sorted into quintiles according to their CAPM beta is monotonically
decreasing as the portfolio beta increases. These findings are at odds with Sharpe (1964) and Lintner (1965)'s CAPM, and are consistent with a flatter security market line (SML) as in Black (1972).

Frazzini and Pedersen (2013) suggest that a flatter SML is observed empirically because investors are constrained in the leverage that they can use. One of the assumptions used in deriving the CAPM is that investors are free to lend or borrow at the risk-free rate, while in practice many investors, such as individuals, pension funds and mutual funds, are constrained in the leverage that they can use. Their model incorporates this friction and predicts that leverage-constrained agents overweight high-beta assets, resulting in lower returns on those assets compared to the predictions of the frictionless CAPM. They propose a trading strategy involving long positions in low-beta assets matched with short positions in high-beta assets, and denote this the "Betting-Against-Beta" factor. Hong and Sraer (2012) instead suggest that when investors disagree about the common factor in firm cash-flows, high beta stocks experience a greater divergence of opinion about their payoffs, and due to short-sales constraints the high beta stocks remain overpriced. It is interesting to note that both explanations for the anomaly require us to relax our assumptions about frictions in the derivation of the CAPM.

I investigate how the Betting-Against-Beta factor performs during subsamples, and find that for the 1990-2012 subsample the anomalous returns on the portfolio disappear. The structural break in the data is coincident with a rise in hedge fund AUM, see for example Fung and Hsieh (1999, Table 1). This is consistent with Frazzini and Pedersen's leverage-based explanation for the anomaly. This emphasises the importance of frictions when it comes to relating fundamentals to security prices. By neglecting the leverage constraint of investors, the classic CAPM overstates the risk-reward offered by high beta stocks.
Chapter 2
Evidence for Limited Arbitrage in Global Equity Markets

Abreu and Brunnermeier (2002) present a model for limited arbitrage in which synchronization risk leads to the persistence of mispricings caused by behavioural traders. Using price-earnings ratios I identify mispricings in global equity markets, and demonstrate that a simple systematic timing strategy outperforms naive attempts to correct the mispricing. The evidence is that sustained deviations from fundamentals occur in equity markets. This study provides empirical evidence to support the modification of AB (2003) to include a parameter controlling preferences towards synchronization risk.
I. Introduction

There are well documented cases where the market prices of financial assets have been observed to deviate from their fundamental values in a sustained manner. Abreu and Brunnermeier (2002, henceforth AB (2002)) and Scheinkman and Xiong (2003) provide some interesting examples. This contrasts with the efficient market hypothesis, which in its semi-strong form states that market prices reflect all publicly available information [Fama (1965)]. Asset price bubbles may be regarded as extreme cases of mispricings, and the recent credit and housing bubbles are a clear reminder of the propensity for such mispricings to occur and their importance to the real economy. Accordingly, there are a number of theories for how deviations between market and fundamental values might arise. Shiller (2000) famously suggested that irrational exuberance arising from "new economy" thinking was the key explanation for the tech bubble of the late 1990s. In Scheinkman and Xiong (2003) heterogeneous beliefs and short sales constraints result in overpricing. Ofek and Richardson (2003) argue that in the presence of short-sale restrictions, the market clearing price is equated to the demand price of the optimistic agents since pessimists' opinions have a limited price impact. Allen et al (2006), DeLong et al (1990a) and Bacchetta and van Wincoop (2006) explore how higher-order beliefs can cause agents to buy at prices that exceed their own fundamental valuations, since they expect to be able to sell to other agents for an even higher price. DeMarzo et al (2008) explore how relative wealth effects can induce herding into risky securities.

In the limits to arbitrage literature, Shleifer and Vishny (1997) describe how rational arbitrageurs who are aware of a mispricing may delay taking a corrective bet because of the risk that the mispricing worsens during their performance evaluation horizon, leading to underperformance and fund outflows due to the agency problem. This would leave them without capital precisely when the arbitrage opportunities are best. In AB (2002) arbitrageurs become sequentially aware that the market price has deviated from fundamental value, but coordination from a sufficient number of arbitrageurs is required to endogenously correct the mispricing. Since arbitrageurs do not know who else is aware of the mispricing, they do not know whether a sufficient mass of their peers are informed to correct a mispricing. As a result they are exposed to synchronization risk if they naively trade to correct the mispricing too early. Synchroniza-
tion risk and the costs of carry induce the informed set of arbitrageurs to delay their corrective trades and implement a timing strategy, thereby prolonging the mispricing.

This study is about mispricings in financial markets as described in AB (2002). Abreu and Brunnermeier (2003, henceforth AB (2003)) apply the principles of synchronization risk to the special case of bubbles. In AB (2003) the incentive for arbitrageurs to time the market is due to the higher growth rate enjoyed during the bubble regime, while in AB (2002) the incentive comes from the costs of carry and loss of diversification that occur when the arbitrageurs take on an arbitrage position. In AB (2003) bubbles exhibit a temporarily higher price growth rate, while in AB (2002) mispricings occur due to a discrete shock to market value. These mechanical differences aside, AB (2003) can be interpreted as a special case of AB (2002). In this chapter I empirically identify episodes where market values have strayed above fundamental values. These episodes are characterised by market returns that are ex ante predictably negative. Because the empirical sample comprises overpricings, it becomes most elegant to use AB (2003) rather than AB (2002) as the theoretical model in this study. Empirical results in Section VII do suggest that the overpricing episodes studied result from a temporarily higher price growth rate, matching the specification of the price path given in AB (2003). But nonetheless the modelling choice comes with the caveat that there are differences between AB (2002) and AB (2003) as noted above.

I examine aggregate equity indices from 47 different countries and identify overpricing episodes using their ex ante cyclically adjusted price-earnings ratios (CAPE). Valuation metrics such as price-earnings ratios are a cornerstone of fundamental stock analysis [Graham and Dodd (1934), Shiller (2000), etc.], and so it is reasonable to assume that they should be regarded as information generators of relevance to market prices. I take the time series of returns observed during the overpricing episodes, and test the optimality of five different trading strategies. Each strategy assumes that the arbitrageur is initially long in the market, and can choose between the market and the local risk-free asset. Three of these strategies delay betting on a return to fundamentals, aiming to profit from the higher price appreciation rates offered during the overpricing regime before timing their exit by following market signals. One of the strategies immediately sells out of the market, and the final strategy remains long in
the market. The null hypothesis is that it is optimal for an arbitrageur to bet on a return to fundamentals immediately following the identification of an overpricing. However a simple market-timing strategy is found to offer robust and statistically significant risk-adjusted returns that are superior to holding the index, or naively selling the index immediately. These results are consistent with AB (2002, 2003), thus corroborating the existence of synchronization risk.

This chapter has a number of similarities with Guenster and Kole (2013, henceforth GK). GK use standard factor models\textsuperscript{4} to identify bubbles in the Fama and French US industry portfolios from 1964 to 2009. They empirically investigate the optimal strategy of a utility-maximising investor who has the choice between a bubbly industry portfolio and the broader market, and find that investors with horizons longer than 6 months should short the asset bubble, but if investors can rebalance their portfolios within 4 months or less it becomes optimal to ride the bubble. There are some key methodological differences when compared to the study documented in this chapter. GK use factor models to identify bubbles in US industries, while I use CAPE ratios to identify overpriced global stock indices. GK give the investor a choice between holding the bubbly industry portfolio or the broader market, while I give the investor a choice between the local risk-free asset and the local overpriced market. Both studies use a Markov-regime switching model to capture the different phases of the overpricing, but while GK have their investor make a probabilistic inference about the existence of the bubble, I assume that the investor becomes certain of the existence of the overpricing, as in AB (2003). GK predict the optimal decision by maximizing the investor’s utility using the distributional parameters of returns, while I calculate the returns that would result from different strategies, and compare the realized utilities to evaluate which choice is optimal. Nonetheless despite their differences, both studies arrive at similar results.

This study also provides empirical evidence to support the modification of AB (2003) to include a parameter controlling preferences towards synchronization risk. As well as deepening our understanding of the mechanics of limited arbitrage, this work therefore aims to clarify the theoretical channel via which policy-makers could improve market efficiency, i.e. by influencing arbitrageurs’ preference towards synchronization risk.

\textsuperscript{4}Such as the CAPM, Fama-French 3-factor model and Carhart (1997)’s 4-factor model.
II. Data Collection

Daily market price observations for aggregate stock market indices were taken from Datastream for a sample of 47 different countries, along with their associated earnings and dividends. The sample includes but is not limited to the G20, Western Europe, South America and the developed East Asian economies. In each case the maximum available time series was selected. While developed markets such as the UK, Germany, Australia and the United States have aggregate earnings data available from 1st January 1973, emerging and developing economies such as Brazil and Pakistan have earnings data that only begins in the early 1990s. The time span of the entire study therefore covers 1st January 1973 till 31st December 2010, but not all countries have data available for the entire period.

Further to this, a local low risk lending rate was obtained from Datastream for each country in the study. The aim here was to obtain a proxy for the local risk-free rate, or the opportunity cost of capital. In the vast majority of cases the rate used was the local short-term interbank rate, but for the following cases the secondary market yield on local short-term government debt was used instead: the United States, Canada, Egypt, Israel, the Netherlands, New Zealand, Philippines, South Africa, Singapore and the UK. Where a choice between the two was available, the rate deemed to contain the lowest risk was chosen. For many countries low risk lending rate data only becomes available after 1st January 1973, but these gaps in the data only coincided with identified mispricing episodes (please see the next section for more details) in two cases, both occurring in the in-sample of Version 2 of the study (more on that later). Please see Appendix 1A for details on how rates were backcasted to fill in these gaps.

Finally, World Bank GDP deflators (also from Datastream) were used to construct price levels for each country in the study. The GDP deflator was used in order to obtain a comparable price level measure for all countries in the study, covering the entire time span of the study, 1st January 1973 till 31st December 2010. Note that the UK for example does not have CPI data until 1988.

III. Identifying Mispricing Episodes
There is an extensive literature that aims to forecast stock market returns using predictive regressions on state variables such as valuation ratios, interest rates and other financial indicators. Welch and Goyal (2008) re-examine the performance of these models and find that most are not significant in-sample, let alone out-of-sample. Their conclusion is that such models would not be useful in helping an investor profitably time the market.

Looking at the S&P Composite Stock Price Index over January 1881 to January 2000, Shiller (2000) retrospectively identifies mispricing episodes as occurring when the aggregate price-earnings ratio reaches a local maximum. Note that Shiller uses the 10-year moving average of real earnings in the denominator of the ratio to correct for the business cycle and other fluctuations. These are known as cyclically adjusted price-earnings ratios (CAPE). Intuitively the idea is that for a given level of fundamentals only a certain range of prices can be supported indefinitely, and a high price-earnings ratio cannot last forever.

A price-earnings ratio can fall for two reasons: either (a) because growth has occurred and earnings have risen to support the higher prices, or (b) because prices have fallen upon recognition that earnings will not rise. Nonetheless Shiller’s ex post local maxima all correspond to acknowledged stock market bubbles. These are the 1901 electricity boom, the 1929 stock market bubble, the 1966 electronics boom and the 2000 tech bubble. In each case the real market value of the S&P Composite Stock Price Index can be seen to drop sharply during the subsequent years. This matches a conventional definition of bubbles as an extended rise in prices followed by a sharp fall, as in AB (2003). More generally the episodes can be viewed as overpricings, where market prices exceed fundamental values as in AB (2002). Either way it seems that periods of growth, where price-earnings ratios are justifiably higher than normal, are sometimes followed by periods of over-optimism resulting in local maxima. One might hypothesize that the local peaks correspond to a stage where market participants have in fact started overreacting to the good news, bidding up prices beyond what is justified by the fundamentals.

However it is necessary to have ex post information on earnings before we can use Shiller’s technique. Without this it is not possible to know whether a local maximum is occurring, or whether the local maximum is going to occur at some later time. In other words, without

20
the ex post earnings information it is not possible to know whether a high price-earnings ratio corresponds to the justified growth phase, or the overpricing phase that sometimes occurs later. Therefore another technique will be necessary to identify mispricing episodes ex ante, as is required of the arbitrageurs to be modelled in this study.

There are other issues too. Shiller’s data on the US covers a time-span of considerably greater length than is available for any of the other 46 countries in the study. Furthermore the price-earnings ratios of different stock markets have different time series characteristics, as one might expect given heterogeneity in the cross-section of markets. A consequence of this is that when identifying mispricings it is not possible to use the same parameters for one country as the next, without somehow controlling for fixed effects. The main approach used to handle these issues is described below in section IV. This is the principal version of the study. For robustness two other versions of the study are also discussed in section VI. To help distinguish between the different versions of the study, let us denote the principal version as Version 1 (V1), and the two robustness checks as Version 2 (V2) and Version 3 (V3). Throughout, the information set of the arbitrageur is the same as that of the econometrician. The arbitrageur sees all countries in the dataset, but only uses those with sufficiently long data series for his econometric analysis. Although he uses international data, it is assumed that he is running funds each with a geographic mandate that forces him into a choice between a local stock market index and a local low risk lending rate. This is a realistic setup since it is typical in the fund management industry for funds to have a specific geographic mandate.

**IV. Method**

The time span of the entire study was divided into the in-sample (henceforth, IS), 1st January 1973 till 31st December 1999, and the out-of-sample (henceforth, OOS), 1st January 2000 till 31st December 2010. Note that 19 countries are consequently excluded from this version of the study altogether, since they lack enough earnings data to have any price-earnings ratios in-sample. Remember we need the 10-year moving average of real earnings to calculate CAPE ratios. This leaves us with 28 countries for V1. Although this is not ideal, the advantages of
For each country in the study a time series of CAPE ratios was calculated. To control for heterogeneity, the quantile of each observed CAPE ratio was calculated with respect to the CAPE ratios observed IS for the country in question. Of the 28 countries included, 16 have more than 4000 daily CAPE ratio observations IS, and the remainder mostly have more than 500 observations, although one country does have only 130 observations. Therefore in general we may be satisfied that there is sufficient IS data for the non-parametric quantile estimation. Hence for each country the CAPE ratio time series was re-expressed as a time series of quantiles of CAPE ratios. A priori it was then assumed that if a country’s CAPE ratio exceeds the 95th quantile for that country, it is possibly in a growth or overpricing phase. To control for growth situations, each country that exceeds its respective 95th quantile was then ranked cross-sectionally using its quantile, and the top 2 countries were deemed to have crossed the threshold into an overpricing episode at the time that they entered the top 2. This point in time is referred to as episode-start. This process for identifying overpricing episodes might seem somewhat arbitrary, but note that the results of the study are robust to variation in the parameters used, and market returns are predictably negative following the identification of an overpricing episode. This systematic analysis using price-earnings ratios therefore succeeds as a robust predictor of market returns.

Previously a decision was made to have a uniform cut-off date between the IS and OOS. This is necessary for the integrity of the cross-sectional comparison. All countries must be in the IS and OOS phases at the same time. At any time each country’s quantile of CAPE is calculated using their IS data, so if one country is IS and another OOS at time $t$, then the cross-sectional comparison of quantiles would be using information from the future, and this would be unacceptable for the OOS tests. Although one fix would be to use only the information up until time $t$ to calculate the quantiles of CAPE at time $t$, and this would ensure that no information from the future is used to make the cross-sectional comparison,
one significant drawback is that it would lead to very poor estimation of quantiles near the start of the time series, and since an IS period is required anyway for the parameterization of the market-timing strategies, this alternative seems like a poor choice. Having a uniform cut-off date between the IS and OOS therefore ensures that the arbitrageurs being modelled in the study do not in any way use information from the future when choosing their positions for the OOS tests.

The cross-sectional ranking technique is appropriate given the openness of the economies in the sample, and their interdependency due to trade. When growth is expected to occur it will tend to get priced into more than one market, each of which will then exhibit relatively high CAPE ratios compared to their own histories. This should not necessarily lead us to believe that overpricings are occurring in all those countries, and indeed the relative position of the countries in a list ranked according to the quantile of their CAPE ratio will not change. Crucially then, just having a historically high CAPE ratio will not necessarily lead to a country being classed as overpriced. In this way the technique controls for cases where market participants are reasonably pricing in some expected future growth. However there will be some markets where participants have over-reacted to the positive signals, and an overpricing episode has begun. These markets will exhibit higher quantiles than even the other countries who are in the growth phase. In other words, the markets near the top of the list are the ones that have a CAPE ratio that is historically high (compared to their own history), but more-so than anyone else (compared to their history), and it is these countries that I identify as exhibiting overpricing due to the over-reaction of market participants. There are consequences to this cross-sectional ranking technique. For a start, it reduces the cross-section from 47 countries to 28. Moreover, in sections V and VI it will become clear that the results of the study vary depending on whether or not the technique is applied.

There are theoretical models that specify stock prices to follow a stochastic process, such as the random walk in Black and Scholes (1973). If stock price processes do have a Brownian component, it would be possible for a country’s quantile of CAPE ratio to cross the overpricing threshold repeatedly during short intervals. To avoid the possibility of having dozens of brief and spurious overpricing episodes, a market is only deemed to have conclusively exited the
overpricing phase when it has been out of the top 2 for 12 months (261 trading days) or more. This point in time is referred to as episode-end. After episode-end has occurred, market participants realize that the point in time 12 months ago when the market last dropped out of the top 2 was indeed the end of the overpricing, referred to as overpricing-end. These definitions are summarized below, along with definitions of two other key events, market-peak and market-bottom:

- **EPISODE-START**: This is the first point in time at which a country’s quantile of CAPE ratio exceeds 95% and is ranked in the top 2 cross-sectionally, following either the beginning of the study or the previous episode-end.

- **OVERPRICING-END**: This is the point at which a country’s quantile of CAPE ratio either drops out of the top 2, or ceases to exceed 95%, and stays below those thresholds for 12 months. Hence, market participants are only aware that overpricing-end has occurred 12 months following the event.

- **EPISODE-END**: This is the point at which a country’s quantile of CAPE ratio has remained below 95% or outside of the top 2 for 12 months. When episode-end occurs, market participants also become aware that overpricing-end occurred 12 months prior.

- **MARKET-PEAK**: This is the point over the interval [episode-start, overpricing-end] at which the nominal market value of the index is highest.

- **MARKET-BOTTOM**: This is the point over the interval [overpricing-end, episode-end], at which the nominal market value of the index is lowest. When it comes to evaluating the optimality of the different strategies, this point in time will be used as the end of the evaluation period so as to give the null hypothesis the best chance of success.

It is also useful to summarise the four key parameters used in the overpricing identification system:

- The time between overpricing-end and episode-end: 261 trading days.
• The threshold for the quantile of CAPE: 95%.

• The growth control, i.e. the maximum number of countries that can cross the threshold into episode-start status at one time: 2.

• The cut-off date between the in-sample and the out-of-sample: 31st December 1999.

Using this system 20 overpricings were identified IS, and 10 OOS. Please see the graphs of the OOS overpricing episodes, included as Figures 1 - 9 at the end of this document. The first dashed line of any pair (red) marks episode-start times, and the second dashed line (green) marks episode-end times. Remember that when evaluating the different trading strategies, we are interested in the interval [episode-start, market-bottom], where market-bottom is the point over the interval [overpricing-end, episode-end] at which the nominal market value of the index is lowest. Inspection confirms that 18/20 of the IS overpricings have market values at market-bottom that are lower than at episode-start, and 10/10 for OOS. In other words the systematic analysis of CAPE ratios reliably predicts negative market returns. Importantly, this method for overpricing identification works ex ante (in contrast to V2, described later). Although the choice of parameters is somewhat arbitrary, please see section VI for a description of the work done on verifying the robustness of the study to variation in these parameters. Moreover, in the context of AB (2003)'s sequential awareness model, episode-start here represents the point in time at which an arbitrageur becomes aware of the overpricing. Note that any imprecision in the overpricing identification method is quite consistent with the model, since it captures how arbitrageurs in AB (2003) do not necessarily become aware of the overpricing when it begins, and when they do become aware of the overpricing they do not know with certainty when it began. The necessary criterion is that the method actually succeeds in identifying overpricing episodes, and this is satisfied since market returns are predictably negative in 28/30 episodes.

Arbitrageurs initially have a portfolio weight of 1 allocated to the stock market index. Although the AB model does not consider dividends (Version 3 looks at this more closely), it is assumed here that while long in the index the flow of dividends is collected in the cash account, which itself earns the corresponding low risk lending rate for that market. At the sell-out time the arbitrageur allocates a portfolio weight of 0 to the market and 1 to the cash
account, which then earns the low risk lending rate indefinitely, or at least until market-bottom occurs. This restriction of the action space to either 1 (fully long in the market), or 0 (neutral), arises from Lemma 1 in the AB (2003) model whereby there are no partial purchases or sell outs. Furthermore the arbitrageurs follow a trigger strategy so that if and when they choose to sell out of the market, they remain out for the remainder of the overpricing episode. This property comes from Corollary 1 in the AB (2003) model. Just to clarify, we use slightly different notation to AB (2003) when describing the action space. Here, 1 denotes a long position in the market, and 0 denotes a neutral position. The decision on when and if to sell out is controlled by one of five automated strategies:

1. Strategy 1 sells out of the index immediately following episode-start, switching the portfolio weight of the index to 0, and allocating a portfolio weight of 1 to the cash account. In the classical theory the selling pressure of trades such as this is theorized to correct mispricings, leading to efficient markets [Fama (1965)]. Since this strategy represents the null hypothesis, a number of choices are made to give it the best chance of success, so as to make any possible rejection of the null more convincing.

2. Strategy 2 remains long in the market indefinitely, and at least until market-bottom. This is a simple buy and hold strategy.

3. Strategy 3 attempts to time the market by remaining long, until the point at which the quantile of CAPE ratio crosses either the upper or lower threshold, \( x_u \) and \( x_l \) respectively. If episode-start was defined simply as the point at which a market crosses its 95th percentile, then the thresholds could be defined as absolute percentile levels. However due to the cross-sectional ranking technique, episode-start can occur for any quantile \( > 0.95 \) depending on the relative positions of the countries in the cross-section. The point at which to sell out must therefore also depend on the quantile of the country at episode-start. Accordingly the threshold is expressible as the proportional change in the quantile compared to the quantile at episode-start. E.g. if a market has a quantile of 0.955 at episode-start, and the threshold \( x_u = 1.025 \), then when the quantile exceeds \( 0.955 \times 1.025 = 0.979 \) the arbitrageur would sell out. If the threshold \( x_u = 1 \), then the
arbitrageur would sell out immediately.

4. Strategy 4 attempts to time the market by using a filter rule. It remains long until the price at time $t$ has dropped by at least $p\%$ compared to the maximum value observed so far over the interval $[\text{episode-start}, t]$.

5. Strategy 5 attempts to time the market by using a moving average rule. It remains long in the market until the current price is at least $y\%$ below the moving average of observed prices, where that moving average is taken over the last $\delta$ days.

Strategy 3 uses a similar fundamentals-based rationale as the overpricing identification methodology. Strategies 4 and 5 follow Brunnermeier & Nagel (2004)'s observation that hedge funds may potentially have been using technical signals to forecast the sentiment of less sophisticated investors during the tech bubble. The parameters $x_u$, $x_l$, $p$, $y$ and $\delta$ are chosen using the IS.

Since a range of market-timing strategies are being used it is appropriate to test for optimality by using Romano & Wolf (2005)'s Stepwise Multiple Testing, which is a generalization of White (2000)'s Reality Check that allows multiple hypotheses to be tested. To briefly recap the Reality Check, let's assume that the test statistic being used to measure outperformance of a candidate model vs. the benchmark is asymptotically normal, i.e. $\bar{f} \rightarrow N(E[f^*], \Omega)$. However in a scenario with $M$ candidate models, the best-performing model’s test statistic is not actually normally distributed, but follows the distribution $\max_{k=1,\ldots,M} \bar{f}_k$. It would therefore not be fair to test the best-performing candidate model using a normal distribution, and indeed White’s test proceeds by constructing the required distribution of extreme values using resampling techniques, so as to provide a fair test that corrects for the data mining issue. Essentially it remains a differences-in-mean test, but which diminishes the significance of any test statistic which suffers from data mining issues. Romano & Wolf (2005) generalizes this by performing multiple rounds of White (2000)'s Reality Check. If no model is found to be significant in the first round of testing, then the procedure stops there. However if there are one or more significant models in the first round of testing, then a second round of testing is performed using extreme values constructed without using those models. If none of the remaining models are found to be significant in the second round, then the procedure stops.
there, et cetera. The interested reader is referred to the relevant papers for more details. For the purpose of this investigation, the test statistics will be:

(i) the difference in discounted sums of log-utility of wealth between the candidate strategy $k$ and the benchmark

\[
\hat{f}_{k,t} = \beta^t \ln W_{k,t} - \beta^t \ln W_{b,t}
\]

\[
\Sigma f_k = \sum_{t=0}^{T} \beta^t \ln W_{k,t} - \sum_{t=0}^{T} \beta^t \ln W_{b,t}
\]

Where $W_{k,t}$ denotes the wealth held at time $t$ by following candidate strategy $k$, and $W_{b,t}$ denotes the wealth held by following the benchmark strategy, which is Strategy 1. Here $\Sigma f_k$ is the test statistic, i.e. the difference in the discounted sum of utilities. Since this sum will actually be taken across a pooled sample of overpricing episodes, it is necessary to reset the wealth level to $W = 1$ and the impatience discount to $\beta = 1$ at the start of each overpricing episode, so that $\Sigma f_k$ actually consists of the sum of the discounted sum of utilities offered by the overpricing episodes, as if each is viewed ex ante. The test statistic captures the difference in ex ante log utility facing an investor if he chooses the alternative strategy rather than the benchmark. Referring to White (2000), it is important that this test statistic is standardized, so that if the entire experiment were rerun the resulting value for $\Sigma f_k$ would be a draw from the same distribution as before, i.e. $N(\Sigma f^*, \Omega)$. The choice of log utility is primarily made in order to meet this requirement. For a brief discussion of the standardization of $\Sigma f_k$, please see Appendix 1B, which also explains why no other utility specifications were used. An annual discount of $\beta_A = 0.98$ is used, corresponding to a daily discount of $\beta_D = 261 \sqrt{0.98}$ since the Datastream time series have on average 261 daily observations per annum.

(ii) the difference in Sharpe ratio between the candidate strategy $k$ and the benchmark.

\[
\bar{f}_k = S R_k - S R_b
\]
Where the Sharpe ratios (SR) are calculated as in Sullivan et al (1999), and it is left to the reader to observe that $\bar{f}_k$ is standardized. Both of these measures succeed in penalizing a strategy for bearing additional risk, which is essential, but the utility-based measure (i) is most successful in this regard since it also penalizes for higher moments such as skewness and kurtosis, as well as penalizing a strategy the earlier it loses money, reflecting the value investors assign to holding wealth over time. It is therefore the higher-priority criterion throughout the study.

To choose parameters ($x_u$, $x_l$, $p$, $y$ and $\delta$) for the market-timing strategies it is therefore appropriate to pick values which maximize criteria (i) for the IS overpricing episodes, since that is the higher priority criterion in the study. For each attempt (an attempt being a specific strategy with a specific candidate parameter choice, applied to all the overpricings from the IS), the ensuing stream of returns that arise from the market index returns, dividends, cash account returns and transaction costs, over the interval [episode-start, market-bottom], for each overpricing episode sequentially, is used to calculate the statistic. Of course the arbitrageurs being modelled have no way of knowing when market-bottom occurs, except in hindsight when episode-end occurs. However they don’t need to be aware of market-bottom when it happens. All that is necessary is the assumption that they follow trigger strategies described by Strategies 1-5. It then becomes an arbitrary choice of the researcher as to when to end the interval, and market-bottom is chosen in order to give Strategy 1 the best chance of success in the OOS tests. Taking Strategy 3 as an example, each candidate parameter pair resulted in a $\Sigma f_k$ statistic. These were used to perform a grid search in order to find the candidate parameter pair that maximized $\Sigma f_k$ IS. Using this method, the thresholds $x_u = 1.05$ and $x_l = 0.75$ were chosen. This process was repeated for Strategies 4 and 5, resulting in parameters $p = 2.5$, $y = 1.5$ and $\delta = 220$. Note that for any given strategy the same parameters apply to every country in the sample.

The completed strategies, with their now-constant parameters chosen using the IS, were then applied to the OOS, whereby each strategy resulted in a single test statistic for $\Sigma f_k$ and a single test statistic for $\bar{f}_k$. In order to execute Romano and Wolf (2005)’s Stepwise Multiple Testing, it was then necessary to block resample the pooled OOS returns resulting from the
strategies, so as to form quantiles of the distribution of extreme events. Please see Appendix 1C for a detailed discussion of the resampling methodology, and the Studentization of the test statistics.

As a robustness check, V1 was repeated in a version which allows short positions, so that if and when the arbitrageurs sell out of the market they switch the portfolio weight of the index to -1, and allocate a portfolio weight of 1 to the cash account. Note that short-selling does not raise any funds for the arbitrageur since the proceeds must be placed as collateral with the broker. It was also assumed that the broker does not mark-to-market the short position until the final period. So even though the price of the index may continue to rise before market-bottom, the study does not assume that the arbitrageur will face collateral calls from the broker until market-bottom occurs, when the profit or loss is booked all in one go. This is to give Strategy 1 the best chance of success. Even if the arbitrageur shorts the index right at the start of the overpricing episode, the collateral calls will not cause them to go bankrupt. Only when market-bottom occurs does the arbitrageur have to cover any short positions, by which time the market is likely to have dropped in value. This is a tremendous advantage for Strategy 1 as the volatility of its returns is drastically limited over the interval [episode-start, market-bottom). Although this is unrealistic, it was done deliberately since Strategy 1 represents the null hypothesis, and we should be quite demanding of any disproof thereof.

V. Results

Please see Table 1 for OOS performance metrics and sample moments for the trading strategies. Strategy 3 has the highest mean return coupled with moderate standard deviation, skew and kurtosis. Strategy 4 has the highest Sharpe ratio, but also has high skewness and the highest kurtosis. Tables 2 and 3 give the test statistics and $p$-values for V1. Each Studentized test statistic is followed by its associated $p$-value in brackets underneath. Note that because the Reality Check $p$-values are corrected for data mining, it is not unusual for a positive test statistic to have a $p$-value > 0.5. This result indicates that the candidate strategy does outperform the
benchmark, albeit not in a statistically significant way given the correction for data mining. An asterisk next to the $p$-value indicates that it was obtained in the first step of the Stepwise Multiple Test, and two asterisks indicate that the $p$-value was obtained in the second step, et cetera. The best $p$-value score for each strategy is recorded. Note that the chosen significance level is 10%, meaning that if none of the remaining test statistics are significant at the 10% level in any given round, the stepwise testing is terminated there.

Table 1: Trading Strategy Performance Metrics for V1.

For each strategy the OOS returns achieved were pooled across all overpricing episodes, and these observations were used to calculate standard moments and performance metrics. Note that for Strategy 2 the case with shorts is equivalent to the case without shorts, since the index is never sold in the buy and hold strategy.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Mean</th>
<th>St. dev</th>
<th>Sharpe ratio</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 no shorts</td>
<td>1.257e-4</td>
<td>1.531e-4</td>
<td>-0.056</td>
<td>-13.924</td>
<td>228.269</td>
</tr>
<tr>
<td>shorts</td>
<td>5.168e-4</td>
<td>0.010</td>
<td>0.037</td>
<td>24.681</td>
<td>672.200</td>
</tr>
<tr>
<td>2 no shorts</td>
<td>-7.236e-4</td>
<td>0.019</td>
<td>-0.045</td>
<td>-0.054</td>
<td>12.513</td>
</tr>
<tr>
<td>shorts</td>
<td>2.248e-4</td>
<td>0.006</td>
<td>0.015</td>
<td>-0.201</td>
<td>47.807</td>
</tr>
<tr>
<td>3 no shorts</td>
<td>6.829e-4</td>
<td>0.013</td>
<td>0.043</td>
<td>16.920</td>
<td>405.078</td>
</tr>
<tr>
<td>shorts</td>
<td>1.939e-4</td>
<td>0.003</td>
<td>0.022</td>
<td>-3.655</td>
<td>264.657</td>
</tr>
<tr>
<td>4 no shorts</td>
<td>6.224e-4</td>
<td>0.011</td>
<td>0.046</td>
<td>20.829</td>
<td>513.941</td>
</tr>
<tr>
<td>shorts</td>
<td>1.126e-4</td>
<td>0.012</td>
<td>-0.002</td>
<td>-1.071</td>
<td>25.033</td>
</tr>
<tr>
<td>5 no shorts</td>
<td>5.180e-4</td>
<td>0.016</td>
<td>0.024</td>
<td>6.346</td>
<td>131.160</td>
</tr>
</tbody>
</table>

Strategy 3 offers significantly higher $\Sigma f_k$ statistics than Strategy 1, both for the cases with and without short positions. Strategy 3 also offers significantly higher $\bar{f}_k$ statistics for the case without shorts, although the significance of the relationship disappears for the case that includes short positions. Table 3 reveals that Strategy 4 performs marginally better against the benchmark than Strategy 3, and therefore according to the $\bar{f}_k$ criterion one should choose the filter rule. However not only is the log utility-based statistic the higher priority criterion, as
Table 2: Stepwise Multiple Hypothesis Tests for V1 Using Log Utility.

For each strategy one test statistic $\Sigma f_k$ is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the $p$-value indicates that it was obtained in the first round of testing, and two asterisks indicates the $p$-value was obtained in the second round, et cetera. The best $p$-value score for each strategy is recorded. $T^{1/3}$ and $T^{3/4}$ indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no shorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>0.1480</td>
<td>1.8246</td>
<td>0.5749</td>
<td>0.5638</td>
</tr>
<tr>
<td></td>
<td>(0.5045**)</td>
<td>(0.0560*)</td>
<td>(0.3230**)</td>
<td>(0.3270**)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>0.1491</td>
<td>1.8651</td>
<td>0.5896</td>
<td>0.5499</td>
</tr>
<tr>
<td></td>
<td>(0.5370**)</td>
<td>(0.0580*)</td>
<td>(0.3465**)</td>
<td>(0.3610**)</td>
</tr>
<tr>
<td>shorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>0.3164</td>
<td>1.8756</td>
<td>0.5569</td>
<td>0.8008</td>
</tr>
<tr>
<td></td>
<td>(0.5400**)</td>
<td>(0.0915*)</td>
<td>(0.4275**)</td>
<td>(0.3290**)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>0.3254</td>
<td>1.9396</td>
<td>0.5894</td>
<td>0.8019</td>
</tr>
<tr>
<td></td>
<td>(0.5555**)</td>
<td>(0.0795*)</td>
<td>(0.4480**)</td>
<td>(0.3540**)</td>
</tr>
</tbody>
</table>

stated above, but furthermore Strategy 3 offers statistically significant outperformance against the benchmark for a greater range of cases, i.e. $\Sigma f_k$ with and without shorts as well as $\bar{f}_k$ without shorts.

Referring to Table 3, the test statistic for Strategy 3 with short sales allowed is actually inferior to that of Strategy 5 with no shorts allowed. This may be surprising, since Table 1 indicates that the Sharpe ratio for Strategy 3 is considerably higher than that for Strategy 5. The difference arises because the benchmark, Strategy 1, performs considerably better in the case with short sales compared to the case with no short sales. Note that Strategy 3 with short sales performs considerably better than Strategy 5 with short sales.

Overall the results suggest that market-timing is the optimal course of action since at least one market-timing strategy, Strategy 3, unequivocally dominates Strategy 1. Although Strategy 4 and Strategy 5 only offer insignificant performance vs. Strategy 1, it remains the case that one strategy, which may be assumed to model the choices that some subset of
Table 3: Stepwise Multiple Hypothesis Tests for V1 Using Sharpe Ratios.

For each strategy one test statistic $\bar{f}_k$ is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the $p$-value indicates that it was obtained in the first round of testing, and two asterisks indicates the $p$-value was obtained in the second round, et cetera. The best $p$-value score for each strategy is recorded. $T^{1/3}$ and $T^{3/4}$ indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no shorts</td>
<td>0.1445</td>
<td>1.5686</td>
<td>1.5701</td>
<td>1.1954</td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>(0.5095**)</td>
<td>(0.0790*)</td>
<td>(0.0785*)</td>
<td>(0.1350**)</td>
</tr>
<tr>
<td></td>
<td>0.1850</td>
<td>1.8836</td>
<td>1.9369</td>
<td>1.5001</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>(0.4240**)</td>
<td>(0.0315*)</td>
<td>(0.0260*)</td>
<td>(0.0815*)</td>
</tr>
<tr>
<td>shorts</td>
<td>-3.6843</td>
<td>0.5228</td>
<td>1.1427</td>
<td>-0.9555</td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>(1.0000*)</td>
<td>(0.5970*)</td>
<td>(0.2745*)</td>
<td>(0.9955*)</td>
</tr>
<tr>
<td></td>
<td>-3.7049</td>
<td>0.5129</td>
<td>1.1706</td>
<td>-0.9174</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>(1.0000*)</td>
<td>(0.5975*)</td>
<td>(0.2495*)</td>
<td>(0.9975*)</td>
</tr>
</tbody>
</table>

arbitrageurs make, does strictly dominate Strategy 1. Please see the section below for an examination of the robustness of this result to variation in the parameters of the overpricing identification system. As a further robustness check the entire study was repeated twice, once using a different method for overpricing identification, and once using a slightly different dataset. These alternate versions are denoted Version 2 and Version 3 respectively. They are also discussed below in section VI.

VI. Robustness Checks

As mentioned in section IV when describing the overpricing identification method, although the four parameters used do result in episodes where the market return is predictably negative, the choice of parameters nonetheless remains somewhat arbitrary. In order to explore the robustness of the main result to changes in these parameters, Version 1 was repeated a number
of times with incremental variations to the parameters so as to map out the boundary at which the main result ceases to hold. The trading strategy parameters \((x_u, x_l, p, y, \delta)\) were held constant throughout this exercise, so that the only variation in the setup came from the four overpricing identification parameters. No new IS optimizations were run. The procedure held three of the overpricing identification parameters constant with the same values used as in section IV, while the fourth parameter was incremented (and later decremented) repeatedly until Strategy 3 ceased to be significant at the 10% level for the log utility measure in any of the four cases: with or without shorts, and for block sizes of \(T^{1/3}\) or \(T^{1/4}\). Only the log utility measure was used because as explained in section IV that is the higher priority criterion in this study. The procedure was then repeated but with a different parameter under variation, and holding the other three constant. This continued until each parameter had been individually subject to variation. In this sense the main result of the study remains robust for the following parameter ranges:

- The time between overpricing-end and episode-end: 225 to 300 trading days.
- The threshold for the quantile of CAPE: 95% to 99%.
- The growth control, i.e. the maximum number of countries that can cross the threshold into episode-start status at one time: 1 to 2.

With regards to the fourth parameter, the cut-off date between IS and OOS, the picture is more complicated. Holding the other three parameters fixed, shifting the cut-off date by a year or so in either direction results in the loss of the statistical significance of the result. To consider why this might be note that increasing the cut-off date decreases the size of the OOS, which diminishes the asymptotic limiting behaviour of the test statistic making the test less reliable. On the other hand, decreasing the cut-off date reduces the amount of information we have for calculating non-parametric quantiles for the CAPE ratios. So moving the cut-off date in either direction has drawbacks that diminish either the technique used for testing significance or the technique used for identifying overpricing episodes. In fact, the reason that

\[^{6}\text{And in fact for values much lower than 95% as well, although decreasing the parameter too far below 95% may start to yield samples of overpricing episodes which do not match the required definition.}\]
the original cut-off date of 31st December 1999 was chosen was so as to obtain an optimum compromise there. However remember that the main result of the study is robust to variation in the other three parameters over quite a wide range, and in fact it happens that there is a combination of parameters which retains the main result whilst also being robust to variations in the cut-off date. If we choose 261 trading days between overpricing-end and episode-end, 95% as the threshold for the quantile of CAPE, and 1 for the growth control, then the main result is robust to cut-off dates at least over the range 31st December 1999 to 18th October 2001.

**Version 2:**

Version 2 (V2) uses an overpricing identification method that is identical to Version 1 (V1) except in two key respects. Firstly, with the aim of including all 47 countries in the test sample the cut-off point between the in-sample (IS) and out-of-sample (OOS) is defined on a country-by-country basis as the point at which half of that country’s CAPE ratio time series has passed. An obvious advantage of this is a larger sample, but a downside is that some of the IS overlap some of the OOS which regrettably diminishes the integrity of the tests in V2. The problem is that the arbitrageurs have access to the entire IS when parameterizing Strategies 3-5, so consequently when they approach the OOS they are effectively using information from the future of some markets (e.g. Brazil) to trade in the present for other markets (e.g. UK), albeit in an indirect way. However given that the future information comes from different markets, it might be suggested that this is not a critical issue, and at the least that the drawbacks are partly offset by the benefits of the increased sample size.

Secondly, to identify overpricing episodes a similar method will be followed as in V1 but without any cross-sectional ranking. Here episode-start corresponds simply to the first point in time when a country’s CAPE ratio reaches its 95% quantile, following either the beginning of the study or the previous episode-end. Similar rules then follow regarding overpricing-end, i.e. it is the point at which a country’s quantile of CAPE ratio ceases to exceed 95% and stays below 95% for 12 months, and episode-end, i.e. it is the point at which a country’s quantile of CAPE ratio has remained below 95% for 12 months. As one might suspect however this
method for identifying overpricings is not perfect because it also picks up plenty of growth episodes, where the nominal market value of the index continues to grow indefinitely in response to earnings growth. In fact only 40/43 of the IS episodes and 13/20 of the OOS episodes have nominal market values that are lower at market-bottom than at episode-start.

To correct for the growth issue all the overpricing episodes where nominal market values are not lower at market-bottom than at episode-start are removed ex post. The interpretation is that in reality arbitrageurs would have access to a range of real time information that they could use to discriminate between overpricing episodes and growth episodes ex ante. Much of this information is not available or feasible to process when backtesting, where we are restricted to mechanical rules. The removal of the growth episodes in hindsight is designed to replicate the ex ante discriminatory power the arbitrageurs have in practice. It is important to note then that for at least two reasons (overlapping IS and OOS, and the need to correct for growth ex post), V2 is not entirely satisfactory. Hence the results obtained with this method are used just as a secondary verification of the results from V1. V2 is a robustness check with caveats.

Using this system 40 episodes were identified IS and 13 OOS. Note that Strategy 3 works in a slightly different way for V2 compared to V1. Here the threshold for an overpricing episode is fixed at the 95% quantile, so the upper and lower thresholds for selling out, $x_u$ and $x_l$, can instead be absolute quantile levels for the CAPE ratio. Following similar IS parameterization methods as used in V1, $x_u = 0.956$ and $x_l = 0.86$ are chosen, meaning that once episode-start passes arbitrageurs following Strategy 3 will sell out if the quantile of CAPE ratio for that country exceeds 0.956 or falls below 0.86. For Strategies 4 and 5 parameters $p = 3$, $y = 1$ and $\delta = 40$ were chosen. The same method was used for OOS tests as in V1. Please see tables 4 and 5 for the results.

In all cases Strategy 3 delivers consistently higher log utility than Strategy 1. Furthermore for the case without short positions, Strategy 3 offers positive statistics for the Sharpe ratio-based test as well. However once data mining is accounted for using White’s Reality Check, none of the results are statistically significant. The results of V2 are far weaker and far less conclusive than the results of V1. To consider why this might be, note that the overpricing identification
Table 4: Stepwise Multiple Hypothesis Tests for V2 Using Log Utility.

For each strategy one test statistic $\Sigma f_k$ is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the $p$-value indicates that it was obtained in the first round of testing, and two asterisks indicates the $p$-value was obtained in the second round, et cetera. The best $p$-value score for each strategy is recorded. $T^{1/3}$ and $T^{3/4}$ indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no shorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>0.5818</td>
<td>0.2217</td>
<td>-2.2176</td>
<td>-1.7948</td>
</tr>
<tr>
<td></td>
<td>(0.6765*)</td>
<td>(0.8290*)</td>
<td>(1.0000*)</td>
<td>(1.0000*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>0.6106</td>
<td>0.2675</td>
<td>-2.0385</td>
<td>-1.6781</td>
</tr>
<tr>
<td></td>
<td>(0.6310*)</td>
<td>(0.7870*)</td>
<td>(1.0000*)</td>
<td>(0.9995*)</td>
</tr>
<tr>
<td>shorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>0.7456</td>
<td>0.2173</td>
<td>-2.0863</td>
<td>-1.6689</td>
</tr>
<tr>
<td></td>
<td>(0.4135*)</td>
<td>(0.6970*)</td>
<td>(0.9990*)</td>
<td>(0.9970*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>0.7964</td>
<td>0.2441</td>
<td>-1.9475</td>
<td>-1.5737</td>
</tr>
<tr>
<td></td>
<td>(0.3640*)</td>
<td>(0.6680*)</td>
<td>(0.9985*)</td>
<td>(0.9970*)</td>
</tr>
</tbody>
</table>

The technique used to select the sample in V2 was in three ways less satisfactory and less realistic than that used in V1. As well as only working ex post and having overlapping IS and OOS, furthermore it generates a sample for which buying and holding the index appears to perform better than selling or market-timing, as indicated by the size and positivity of the test statistic for Strategy 2 - although this is not statistically significant either. This suggests that not only is the overpricing identification method used in V2 unrealistic, but it also does not work particularly well and perhaps fails to fully correct for the growth issue. It seems reasonable to assume that selling out of an overpricing, or timing the market, should dominate holding on to the asset while it falls back to fundamental levels. Hence, it would appear that the technique used in V2 does not identify overpricings particularly well, and accordingly less weight should be assigned to the results.

This difference between the results of V1 and V2 suggest that the results of the study are sensitive to the use of the cross-sectional ranking technique, and/or to the number of countries.
Table 5: Stepwise Multiple Hypothesis Tests for V2 Using Sharpe Ratios.

For each strategy one test statistic \( \bar{f}_k \) is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the \( p \)-value indicates that it was obtained in the first round of testing, and two asterisks indicates the \( p \)-value was obtained in the second round, et cetera. The best \( p \)-value score for each strategy is recorded. \( T^{1/3} \) and \( T^{1/4} \) indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no shorts</td>
<td>(-0.3276) ( (0.7430^*) )</td>
<td>(0.2461) ( (0.5075^*) )</td>
<td>(0.2134) ( (0.5225^*) )</td>
<td>(0.0803) ( (0.5845^*) )</td>
</tr>
<tr>
<td></td>
<td>(-0.4279) ( (0.7840^*) )</td>
<td>(0.3231) ( (0.4980^*) )</td>
<td>(0.2802) ( (0.5165^*) )</td>
<td>(0.1049) ( (0.5905^*) )</td>
</tr>
<tr>
<td>shorts</td>
<td>(-3.9376) ( (1.0000^*) )</td>
<td>(-0.8495) ( (0.9985^*) )</td>
<td>(-1.9091) ( (1.0000^*) )</td>
<td>(-2.3739) ( (1.0000^*) )</td>
</tr>
<tr>
<td></td>
<td>(-4.1335) ( (1.0000^*) )</td>
<td>(-0.9553) ( (1.0000^*) )</td>
<td>(-1.7094) ( (1.0000^*) )</td>
<td>(-2.1738) ( (1.0000^*) )</td>
</tr>
</tbody>
</table>

included in the sample. As discussed above, the overpricing identification technique used in V1 is in three ways more satisfactory than V2, and if the cross-sectional ranking technique is necessary to successfully identify overpricings then it will consequently affect the result. However there also remains the possibility that the results are dependent on the method and sample.”

Version 3:

In their model AB assume that the asset in question pays no dividends, assuming for example that the earnings on the stock are reinvested so as to give the holder capital gains as a return. This results in a simplified payoff function for their model. However Versions 1 and 2 of this study look at real stock indices which pay out earnings as dividends. In order to improve the mapping between the empirical work and the theoretical model, Version 3 repeats the study.
but using the Total Return Index (TRI) as the price. New CAPE ratios were calculated, new quantiles estimated, and new overpricing episodes were identified, all using the same techniques as used in Version 1. The same four overpricing identification parameters were used as well.

Using this system 16 overpricings were identified IS and 9 OOS, where 14/16 and 8/9 have nominal market values that are lower at market-bottom compared to episode-start. The trading strategy parameters chosen for V3 are as follows: $x_u = 1.05$, $x_l = 0.75$, $p = 2.5$, $y = 1.5$ and $\delta = 220$. Please see Tables 6 and 7 for the OOS results.

Table 6: Stepwise Multiple Hypothesis Tests for V3 Using Log Utility.

For each strategy one test statistic $\Sigma f_k$ is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the $p$-value indicates that it was obtained in the first round of testing, and two asterisks indicates the $p$-value was obtained in the second round, et cetera. The best $p$-value score for each strategy is recorded. $T^{1/3}$ and $T^{1/4}$ indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no shorts</td>
<td>-0.8839</td>
<td>0.4671</td>
<td>-1.3121</td>
<td>-1.1021</td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>(0.9830*)</td>
<td>(0.6280*)</td>
<td>(0.9965*)</td>
<td>(0.9930*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>-0.8889</td>
<td>0.4671</td>
<td>-1.2702</td>
<td>-1.0236</td>
</tr>
<tr>
<td></td>
<td>(0.9860*)</td>
<td>(0.6165*)</td>
<td>(0.9970*)</td>
<td>(0.9900*)</td>
</tr>
<tr>
<td>shorts</td>
<td>-0.7103</td>
<td>0.4803</td>
<td>-1.1827</td>
<td>-1.0593</td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>(0.9410*)</td>
<td>(0.4975*)</td>
<td>(0.9830*)</td>
<td>(0.9790*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>-0.7297</td>
<td>0.4752</td>
<td>-1.1619</td>
<td>-0.9950</td>
</tr>
<tr>
<td></td>
<td>(0.9460*)</td>
<td>(0.4995*)</td>
<td>(0.9830*)</td>
<td>(0.9700*)</td>
</tr>
</tbody>
</table>

The results are similar to V2. Strategy 3 delivers consistently higher log utility than Strategy 1 in all cases. Furthermore for the case without short positions, Strategy 3 offers positive statistics for the Sharpe ratio test. However once data mining is accounted for using White’s Reality Check, none of the results are statistically significant. Of course a crucial flaw with V3 is that market participants do not really regard the TRI as the price. Hence they do not
Table 7: Stepwise Multiple Hypothesis Tests for V3 Using Sharpe Ratios.

For each strategy one test statistic $\bar{f}_k$ is calculated, and the statistics were Studentized as explained in Appendix 1C. Romano & Wolf (2005) Stepwise Multiple Testing was then used to assess the significance of the Studentized statistics. Note that in this context an asterisk does not indicate significance. An asterisk next to the $p$-value indicates that it was obtained in the first round of testing, and two asterisks indicates the $p$-value was obtained in the second round, et cetera. The best $p$-value score for each strategy is recorded. $T^{1/3}$ and $T^{1/4}$ indicate the mean block size used for the resampling.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
<th>Strategy 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>no shorts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>-0.0031</td>
<td>1.1596</td>
<td>0.8162</td>
<td>0.9498</td>
</tr>
<tr>
<td></td>
<td>(0.6525*)</td>
<td>(0.1755*)</td>
<td>(0.2935*)</td>
<td>(0.2510*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>-0.0038</td>
<td>1.4438</td>
<td>0.9997</td>
<td>1.1893</td>
</tr>
<tr>
<td></td>
<td>(0.6945*)</td>
<td>(0.1215*)</td>
<td>(0.2550*)</td>
<td>(0.1900*)</td>
</tr>
<tr>
<td><strong>shorts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^{1/3}$</td>
<td>-4.1596</td>
<td>-0.9200</td>
<td>-1.2794</td>
<td>-2.2645</td>
</tr>
<tr>
<td></td>
<td>(1.0000*)</td>
<td>(0.9985*)</td>
<td>(0.9995*)</td>
<td>(1.0000*)</td>
</tr>
<tr>
<td>$T^{1/4}$</td>
<td>-4.1716</td>
<td>-0.9197</td>
<td>-1.2267</td>
<td>-2.0656</td>
</tr>
<tr>
<td></td>
<td>(1.0000*)</td>
<td>(0.9975*)</td>
<td>(0.9995*)</td>
<td>(1.0000*)</td>
</tr>
</tbody>
</table>

respond to TRI or TRI-earnings information in the same way as they would respond to price or price-earnings information. As a consequence V3 is artificial in a way which could be deemed important to the theoretical mechanisms underlying the study, and accordingly less weight should be assigned to the results. V3 is much like V2, merely a secondary verification of the main result from V1. V3 is a robustness check with strong caveats. However V3 might become more useful later on in the study.

V1 offers strong evidence that market-timing is the optimal course of action, since the strategy that immediately sells out of the overpricings is unequivocally dominated by the most successful market-timing strategy. As explained above, V1 is the strongest and most realistic version of the study and it should take priority in guiding our conclusions. However it remains the case that the positivity of the test statistics for Strategy 3 in V2 and V3 serves to further corroborate the conclusion. Given this evidence for the existence of synchronization risk, what
follows is empirical support for the modification of AB (2003) to include a parameter controlling preferences towards synchronization risk.

**VII. Synchronization Risk Parameter**

As discussed in Section I it is more appropriate to use AB (2003) rather than AB (2002) as the model for the empirical work in this study. I proceed first by quickly recapping the relevant parts of the theoretical model from AB (2003). The rational arbitrageurs are assumed to be long in the market until the time at which they sell out, denoted \( t \). They obtain the following utility from selling out of the overpricing at time \( t \). It is just the expected payoff:

\[
\int_{t_i}^{t} e^{-rt}(1 - \beta(s - T^{s-1}(s)))p(s)\pi(s|t_i)ds + e^{-rt}p(t)(1 - \Pi(t|t_i)) - c
\]

List of definitions:

- \( t_i \) = the time at which arbitrageur \( i \) became aware of the existence of the overpricing. In this sequential awareness model, this time uniquely identifies the arbitrageurs.

- \( \beta(s - T^{s-1}(s)) \) is the overpricing component in the asset price at time \( s \). Following the correction, only the fraction \((1 - \beta(s - T^{s-1}(s)))\) of the asset value remains.

- \( p(s) \) is the asset price at time \( s \), for the overpricing trajectory. So the post-correction asset price at time \( s \) is given by \((1 - \beta(s - T^{s-1}(s)))p(s)\)

- \( \pi(s|t_i) \) is the PDF that arbitrageur \( i \) assigns to the overpricing collapsing at time \( s \).

- \( \Pi(t|t_i) \) is the CDF that arbitrageur \( i \) assigns to the overpricing collapsing at some point prior to \( t \).

- \( r \) and \( c \) represent the discount rate and transaction costs respectively.

The optimal choice is derived by maximizing the expected payoff with respect to \( t \). This is done by differentiating the payoff function with respect to \( t \) and equating to zero, which results in the sell-out condition denoted Lemma 7 in AB:
If arbitrageur $t_i$'s "subjective hazard rate" (the left-hand side of the inequality below) is smaller than the "cost-benefit ratio" (the right-hand side of the inequality below), then trader $t_i$ will choose to hold the maximum long position at $t$.

$$\frac{\pi(t|t_i)}{1 - \Pi(t|t_i)} < \frac{g - r}{\beta(t - T^*-1(t))}$$

Note that $g$ is defined as the asset price growth rate during the overpricing phase. Conversely, if arbitrageur $t_i$'s "subjective hazard rate" is larger than the "cost-benefit ratio", then trader $t_i$ will choose to hold the maximum short position at $t$.

$$\frac{\pi(t|t_i)}{1 - \Pi(t|t_i)} > \frac{g - r}{\beta(t - T^*-1(t))}$$

The above summarizes AB (2003) as it currently stands. Note that the agent's decision on whether to go long or short on the overpricing is dependent only on the objective parameters on the right-hand side ($g$, $r$ and $\beta$), and the subjective probabilities on the left-hand side. No risk preferences come into play in the decision. The agent has no preferences regarding synchronization risk. The aim of this part of the study is to advance on this, to support the modification of AB (2003) to include a parameter controlling preferences towards synchronization risk.

Let's start by conjecturing a new parameter, $\theta$, which controls preferences towards synchronization risk. It is better to do this than simply inserting a well-known utility function (e.g. CRRA) into the payoff function, because the parameter in a concave utility function would capture many things besides preferences towards synchronization risk, and the aim here is to identify these preferences in as clean and clear a way as possible. Furthermore referring to AB (2003), if a concave utility function were used the payoff function would cease to be a linear function of the quantity of stock held. In this case Lemma 1 and Corollary 1 would cease to hold, and the model would lose the property whereby an agent is 100% "in" the market until the exact point when the cost-benefit ratio is exceeded by the subjective hazard rate, at which stage the agent is 100% "out" and stays "out" thereafter. The model would then require a
recursive payoff function, drastically increasing the complexity of the model and reducing the clarity, but without a corresponding increase in insight. Specifying a new utility function for the agent using this parameter (previously this was just the expected payoff):

\[ \int_{t_i}^{t} e^{-rs} \theta (1 - \beta (s - T^{s-1}(s))) p(s) \pi(s|t_i) ds + e^{-rt} p(t)(1 - \Pi(t|t_i)) - c \]

Intuitively \( \theta \) is analogous to the impatience parameter used in the consumption-based derivation of the discrete-time asset pricing model. The more impatient the agent is, the smaller the impatience parameter is, because the agent weights downwards the utility of second-period consumption since it is less important to him. Here, the less averse the agent is to synchronization risk the smaller \( \theta \) is, because the agent weights upwards the payoffs from states where he sells out prior to the overpricing collapsing.

Differentiating the new expected payoff function with respect to \( t \) results in a modified version of Lemma 7, denoted Lemma 7b:

\[ \frac{\pi(t|t_i)}{(1 - \Pi(t|t_i))} < \frac{g - r}{[1 - \theta (1 - \beta (t - T^{s-1}(t)))]} \]

implies that the agent chooses to hold the maximum long position, and:

\[ \frac{\pi(t|t_i)}{(1 - \Pi(t|t_i))} > \frac{g - r}{[1 - \theta (1 - \beta (t - T^{s-1}(t)))]} \]

conversely implies that the agent chooses to hold the maximum short position. Note that the smaller \( \theta \) is, the more likely it is that the left-hand side will dominate the right-hand side, and the agent will choose to hold the maximum short position. This is consistent with an elevated preference for synchronization risk. The aim of what follows is to estimate values for the \( \theta \) parameter for the different strategies, so as to verify whether the parameter varies in a meaningful way given the behaviour of the arbitrageurs. All the strategies described in Section IV work by indicating, at each point in time, whether to go long on the market or short on the market. So each strategy generates a time series of recommendations on whether
to go long or short in the market. These recommendations can be reinterpreted via Lemma 7b. Each strategy, for each overpricing, gives a sequence of inequalities based on Lemma 7b. Provided that the strategy at some point switches from the recommendation of a long position to the recommendation of a short position, the sequence of inequalities will provide an upper and lower bound estimate for $\theta$, for that strategy, and for that overpricing episode. We can denote these upper and lower bounds $\theta^u$ and $\theta^l$ respectively. Referring to AB (2003), $\beta_t$ can be expressed as a function of $g$, $r$ and time. A similar formula was used here, although given the discrete nature of the data a discrete analogue was applied to this task:

$$\beta_t = 1 - \frac{(1 + r)^t}{(1 + g)^t}$$

Hence obtaining numerical values for the upper and lower bounds on $\theta$ only requires estimates of $g$, $r$ and $\pi(t|t_i)/(1 - \Pi(t|t_i))$ for each market, where the last quantity is just the subjective hazard rate, which can be denoted $h$. To estimate these quantities for each market that experiences an overpricing in the full-sample (the full-sample of overpricing episodes is used to aid the convergence of the estimator), a Markov-Switching model is applied to a discrete-time version of the asset price growth model from AB:

$$\frac{P_t - P_{t-1}}{P_{t-1}} = \mu_s t + \sigma_s \epsilon_t$$

Where $P_t$ is the asset price at time $t$, $\mu_s$ is the drift parameter for state $s_t$, $\sigma_s$ is the volatility parameter and $\epsilon_t$ is a white noise process. The relevant states underlying the model are normal (N), overpricing (OP) and correction (C) as follows:

- $s_t = N$ iff $t \in (\text{market-bottom, episode-start})$
- $s_t = \text{OP}$ iff $t \in (\text{episode-start, market-peak})$
- $s_t = \text{C}$ iff $t \in (\text{market-peak, market-bottom})$

Hence:

$$\mu_N = r$$
$$\mu_B = g$$
Importantly note that in hindsight the state can be known with certainty, by cross-referencing the market’s nominal value and CAPE ratio time series with the definitions for episode-start, market-peak, market-bottom, et cetera, given in section IV. Hence estimates for the parameters of the model can be made conditioning on the fact that in hindsight the state at any time is known with certainty. This is a trivial case of the Markov-Switching model, but which usefully provides a well-known procedure for estimating the required quantities $r$, $g$ and $h$:

- $r_i = \text{sample-mean of observed returns that occur during the normal regime for market } i$.
- $g_i = \text{sample-mean of observed returns that occur during the overpricing regime for market } i$.
- $h_i = \text{number of times the transition is made from state OP to state C, as a fraction of the total number of periods spent in state OP, for market } i$.

Note that the $g_i$ parameters estimated are consistently higher than the $r_i$ parameters for the cross-section of overpricing episodes, and this confirms that the overpricing episodes identified do offer a higher price growth rate as specified in AB (2003). Applying this procedure to the full-sample of overpricing episodes (30 episodes for V1, 53 for V2 and 25 for V3) generates a $(\theta^l, \theta^u)$ pair for each strategy, for each episode. What we are interested in here is whether the parameter significantly differs in a meaningful way for the different strategies. Specifically what we want to know is whether $\theta$ values for Strategy 1 are significantly lower than those for Strategies 3-5. This would justify a modified version of AB (2003) which includes a parameter controlling preferences towards synchronization risk. As well as deepening our understanding of the mechanics of limited arbitrage, this will also identify the theoretical channel via which policy-makers could influence arbitrageurs to inhibit mispricings.

To do this a $t$-stat was derived to assess whether the $\theta^l$ values for each of the candidate market-timing strategies are statistically greater than the $\theta^u$ from Strategy 1, corresponding to an increased level of synchronization risk aversion from all the market-timers. Since we are not just looking at the best-performing strategy, we are not focussing on the extreme values of the distribution, and consequently no data mining correction is required and a simple t-test
is appropriate. The derivation of the statistic is in Appendix 1D. Please see Table 8 for the results of these hypothesis tests for V1, V2 and V3. Each line gives the $t$-stat and one-tailed $p$-value on the test for whether the indicated market-timing strategy’s $\theta^l$ is larger than the $\theta^u$ implied by Strategy 1, across the full-sample of mispricing episodes. Note that since we are only interested in the sell-out time, and not the returns or wealth generated by the trades, it is irrelevant whether or not short positions are allowed.

Table 8: Differences-in-Mean T-tests for the Synchronization Risk Preference Parameter. For each strategy and each overpricing episode a parameter pair ($\theta^l$, $\theta^u$) is estimated from the data. The $t$-statistic and one-sided $p$-values indicate to what extent the $\theta^l$ values for the candidate market-timing strategies are statistically greater than the $\theta^u$ values from Strategy 1, which sells out of the overpricing immediately following detection. Higher values of $\theta$ correspond to lower appetite for synchronization risk. Please see Appendix 1D for the derivation of the $t$-statistic.

<table>
<thead>
<tr>
<th>Version</th>
<th>Strategy</th>
<th>$t$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>S3</td>
<td>1.1099</td>
<td>0.1360</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>2.8030</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>3.3342</td>
<td>0.0008</td>
</tr>
<tr>
<td>V2</td>
<td>S3</td>
<td>1.2206</td>
<td>0.1127</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>4.3003</td>
<td>1.95e-5</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>3.7548</td>
<td>1.44e-4</td>
</tr>
<tr>
<td>V3</td>
<td>S3</td>
<td>1.0921</td>
<td>0.1403</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>3.3105</td>
<td>8.93e-4</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>1.1169</td>
<td>0.1350</td>
</tr>
</tbody>
</table>

The $t$-stats would suggest that the market-timing strategies indeed display a higher aversion to synchronization risk than the overpricing-attacking strategy (Strategy 1). This is logical since by their nature they will remain in the market for longer than Strategy 1. Remember that Strategy 3 also unequivocally dominates the naive overpricing-attacking strategy. The dominant strategy therefore involves less appetite for synchronization risk, manifesting in a
higher $\theta$ parameter. It is disappointing to note that the $p$-value here is only on the threshold of significance at the 10% level. This might have occurred due to the strictness of the test, which compares $\theta^u$ vs. $\theta^l$. Nonetheless, at 0.1360, 0.1127 and 0.1403 for V1, V2 and V3 respectively, the $p$-values for Strategy 3 are convincingly low.

VIII. Conclusion

The efficient market hypothesis suggests that arbitrageurs prevent mispricings by taking corrective positions when they are identified. In this study I provide evidence that when an overpricing is identified the optimal course of action may not be to take a corrective position, but instead to remain long in the market so as to profit from some of the growth, before a carefully timed exit from the market. This evidence goes some way to corroborate AB (2002, 2003)’s model, wherein synchronization risk induces arbitrageurs to delay taking corrective positions immediately. In this chapter I suggest that preferences towards synchronization risk can be captured by the newly proposed parameter, $\theta$, which advances on AB (2002, 2003) to identify the channel via which policy-makers could influence arbitrageurs to choose synchronization risk, thus encouraging the endogenous collapse of mispricings.

If policy-makers could diminish the $\theta$ of arbitrageurs, then whenever faced with the opportunity to engage in a market-timing activity it would be more likely that investors would refrain from doing so. Instead they would be more likely to take their chances with the synchronization risk entailed in taking a corrective position against the mispricing. This in itself would reduce the level of synchronization risk, and since all the agents would be aware of this, a new self-reinforcing equilibrium would arise where agents sell out of the mispricing early, bringing the endogenous collapse time closer to the present. In this way, some mild regulation could incentivise arbitrageurs to suppress mispricings in a market that remains a free price system. Further research could explore remuneration structures that optimally incentivise managers so as to overcome the principal-agent problem, while at the same time minimizing the negative externality of sustained mispricings.
It is an assumption of AB (2002, 2003) that a sufficient mass of trades by arbitrageurs can cause the endogenous collapse of a mispricing. However it remains moot whether there is a causal link between hedge fund holdings and price changes. For example Brunnermeier & Nagel (2004)’s sample of hedge funds held only 0.3% of outstanding aggregate equity. However as documented by Fung and Hsieh (1997), hedge funds employ dynamic trading strategies, which suggests that at any given moment a disproportionate fraction of the overall volume of trades will come from them. Thus despite their low holdings as a proportion of total market capitalization, their trading actions may nonetheless potentially be a significant causal factor in the determination of equilibrium prices. It would be ideal if future research could attempt to establish conclusively what causal link, if any, exists between hedge fund holdings and prices.

\[7\] Whose aggregate stock holdings captured about one fifth of the total assets under management in the hedge fund industry at the time.
Appendix 1A: Backcasting of Interest Rates

According to the overpricing identification method used in Version 2, Belgium’s stock market entered one overpricing episode on 3rd July 1987 and another on 25th May 1989, and furthermore Denmark entered an overpricing episode on 18th November 1983. Unfortunately data on Belgium’s one month interbank rate only becomes available on 1st October 1989, and Denmark’s one month interbank rate only becomes available on 1st July 1988. However data on the yield on 5-year Belgian government bonds and Denmark’s central bank’s discount rate, is available going back beyond the 1980s. These rates, which themselves would not be suitable as a proxy for the local risk-free rate, were used to backcast low risk lending rates using a simple one factor regression model of the following form:

\[ y_t = \mu + \beta \Psi_t + \epsilon_t \]

Where:

\( y_t \) = Denmark or Belgium’s one month interbank rate, respectively.

\( \Psi_t \) = Denmark central bank’s discount rate, or the yield on 5-year Belgian government bonds, respectively.

The parameters of these two models were estimated using the overlap between the \( y_t \) and \( \Psi_t \) observations, and thus where required the missing one month interbank rates were substituted using \( \hat{y}_t \) predictions.

---

As an exception, this data item was taken from the Global Insight database.
Appendix 1B: Standardization of $\Sigma f_k$

White (2000, pg 1101) states that his method requires a standardized test statistic with a continuous limiting distribution. These criteria are met by the Sharpe ratio statistic $\bar{f}_k$, see for example Sullivan et al (1999). However it is less obvious when it comes to the statistic based on time-summed discounted utility of wealth:

$$
\Sigma f_k = \sum_{t=0}^{T} \beta^t U(W_{k,t}) - \sum_{t=0}^{T} \beta^t U(W_{b,t})
$$

This is the quantity which, appropriately standardized, needs to have a continuous limiting distribution. Since returns are approximately normal, wealth is lognormal, utility of wealth will just be a monotonic transformation of a lognormal random variable, and hence sums thereof may take any value within a subset of the real line, thus meeting the requirements for a continuous limiting distribution. For the statistic to be standardized, we must believe that successive repeats of the entire experiment will yield values for $\Sigma f_k$ that come from the same distribution, i.e. $N(\Sigma f^*, \Omega)$. Unfortunately the analytical distribution of sums of lognormal random variables is not currently known. Although approximation methods are available, see for example Wu et al (2005), it is desirable to avoid making approximations where possible. One elegant solution is to choose a log utility function, which has the effect of converting $\Sigma f_k$ into the sum of two sequences of normal random variables, provided of course with the assumption that returns are normally distributed. Let’s assume an initial wealth level of $W_{k,0} = 1$. This would be the stream of wealth accruing for Strategy $k$:

$$
W_{k,0} + W_{k,1} + W_{k,2} + W_{k,3} = 1 + e^{r_1} + e^{r_1+r_2} + e^{r_1+r_2+r_3}
$$

In practice discrete compounding is used to calculate the streams of wealth in this study, so as to match the frequency of the data. Since continuous compounding is equivalent to discrete compounding when the rate is chosen correctly, it remains valid to use continuous compounding in this derivation. Summing the discounted log utility of the wealth stream:
\[ \log W_{k,0} + \beta \log W_{k,1} + \beta^2 \log W_{k,2} + \beta^3 \log W_{k,3} + \ldots = \beta r_1 + \beta^2 (r_1 + r_2) + \beta^3 (r_1 + r_2 + r_3) + \ldots \]

If the reader is willing to assume that returns are i.i.d. then in a straightforward way the distribution for \( \Sigma f_k \) will be the same for every repeat of the experiment. Of course it is conventional to acknowledge that returns follow a covariance stationary process. Provided with this, and a long enough time dimension for the sample, the same means and variances will tend to appear during each repeat of the experiment, and get summed together so that the distributional parameters of \( \Sigma f_k \) converge to approximately the same values on each repeat of the experiment. One complication is that returns which appear early on during an overpricing episode will get weighted upwards by time, as they appear in every following wealth observation and receive less of an impatience discount. Hence the larger the number of episodes used in the test, the more reasonable it is to assume that \( \Sigma f_k \) will be standardized, since there will be a greater number of corresponding "early" values coming from other episodes, which will receive similar time and impatience discount weightings. In this case, the early values from the first overpricing episode will not have a disproportionate effect on the distribution of \( \Sigma f_k \) for the experiment, but will be offset and averaged-out by the other episodes, leading to a distribution for \( \Sigma f_k \) that is approximately the same on each repeat of the experiment - in other words, \( \Sigma f_k \) will be standardized.
Appendix 1C: Resampling

When it comes to resampling a stationary time series, there is a well-defined procedure due to Politis and Romano (1994a, b) which samples blocks of observations thus capturing the dependence of the time series data. One uses blocks of random length, distributed according to the geometric distribution with mean length $b$. The choices for $b$ are explained later in this section. The number of resamples to use for each test was set at $N = 2000$. The resampled returns were then used to calculate time series of utility of wealth values to arrive at $\Sigma f_{j,k}$, and Sharpe ratio values to arrive at $\bar{f}_{j,k}$, for $j = 1, \ldots, N$. In the former case there is a necessary extra step to the implementation compared to White (2000): the wealth and impatience discounts are reset to unity whenever a new overpricing episode begins. For example if the first overpricing episode in the actual OOS corresponds to 200 return observations for each strategy, then the 201st resampled vector of returns would be classed as applying to the second period of the second overpricing episode, where wealth has been reset to $W_{j,1} = 1$ for the first period of the second overpricing episode. In this way the resampled returns are used to calculate a sum of the discounted sum of utilities offered by the overpricing episodes, as if viewed ex ante.

As suggested by Romano and Wolf (2005) the test statistics were then Studentized, so as to obtain a higher powered test for several reasonable definitions of power. There are additional subtle improvements that accrue from Studentization, and the interested reader is referred to Romano and Wolf’s paper for more details. Studentization requires an estimate of the standard error of the test statistics, i.e. $se(\Sigma f_k)$ and $se(\bar{f}_k)$. The most feasible option simply takes the standard deviation of the $\{\Sigma f_{j,k}\}_{j=1}^{2000}$ and $\{\bar{f}_{j,k}\}_{j=1}^{2000}$ values obtained from the resampling, thus arriving at values of $se(\Sigma f_k)$ and $se(\bar{f}_k)$ for each model $k$. This methodology for estimating the dispersion of the test statistics appeals to the same rationale as that underpinning the construction of the extreme value quantiles in White’s test, and importantly it uses the same assumptions. Therefore it is most parsimonious to use these estimates to Studentize the test statistics, both for the actual data and the resamples. This is in preference to alternatives involving delta methods and HAC covariance matrix estimation, all of which entail additional

---

9One return observation for each strategy, each corresponding to the same point in time.
assumptions and approximations that it is more elegant to avoid.

Following Hall et al (1995), two repeats of the tests were carried out using block sizes equal to $T^{\frac{3}{4}}$ and $T^{\frac{1}{2}}$ respectively, where $T$ represents the time dimension of the pooled time series of return observations from all overpricing episodes. The first choice minimizes the MSE of the bootstrapped estimate of the standard error, i.e. $se(\Sigma f_k)$ or $se(\bar{f}_k)$. The second choice minimizes the MSE of the estimate of the one-sided distribution we are constructing using the max values from each resample. Refer to Hall et al equations (2.3) and (2.5) respectively, to understand clearly what is being minimized with these choices. Ideally we would minimize both quantities, but since this is not possible the next best alternative is to do one version where one is minimized, and another version where the other is minimized.
Appendix 1D: Derivation of the $t$-statistic

Denote the theta from one strategy as $Y_i$, and the theta from another strategy as $Z_i$, where both arise from an unknown random distribution and $i$ indexes according to overpricing episode:

$$Y_i \rightarrow D(\mu, \sigma^2_Y); \quad Z_i \rightarrow D(\mu, \sigma^2_Z)$$

Each strategy is applied to the same full-sample of overpricing episodes, so the number of observations in the two samples is the same: $N$. We want to compare the sample mean of $\{Y_i\}_{i=1}^N$ to the sample mean of $\{Z_i\}_{i=1}^N$, but without assuming that they necessarily follow the same distribution. It is necessary to allow for covariant observations since the strategies are being applied to the same full-sample of overpricing episodes. The null hypothesis for the statistical test is that both random variables have the same population mean. Subtracting $Z_i$ from $Y_i$, we know from the basic properties of expectations and variances:

$$Y_i - Z_i \rightarrow D(0, \sigma^2_Y + \sigma^2_Z - 2\sigma_{Y,Z})$$

The sequence $\{Y_i - Z_i\}_{i=1}^N$ is assumed i.i.d., and summed to obtain an approximately Normal distribution via the Central Limit Theorem:

$$\sum_{i=1}^N (Y_i - Z_i) \overset{L}{\rightarrow} N(0, N\sigma^2_Y + N\sigma^2_Z - 2N\sigma_{Y,Z})$$

$$\sqrt{N}(\bar{Y} - \bar{Z}) \overset{L}{\rightarrow} N(0, \sigma^2_Y + \sigma^2_Z - 2\sigma_{Y,Z})$$

$$\frac{\sqrt{N}(\bar{Y} - \bar{Z})}{\sqrt{\hat{\sigma}^2_Y + \hat{\sigma}^2_Z - 2\hat{\sigma}_{Y,Z}}} \overset{L}{\rightarrow} St(0, 1, 2N - 2)$$

Since each strategy is applied to the same full-sample of overpricing episodes, we are running a more controlled and hence more powerful comparison test than if each strategy was applied to a different sample of episodes. The covariance term therefore tends to diminish the standard error, resulting in a larger $t$-statistic.
Chapter 3
Distress Risk in the Size and Value Factors

Stock loadings on the Fama-French portfolios are linked to distress risk at the firm-level. Although factor model betas are estimated with error, I use the SIMEX method to correct for this measurement error issue and conclusively link betas to distress risk. Furthermore I use support vector machines to examine the relationship in a nonlinear setting, and find the surprising result that high magnitudes of beta are linked to distress risk. These findings have implications for the interpretation of the size and value factors, and their relationship to distress risk and economic fundamentals.
I. Introduction

The Fama and French (1996, henceforth FF) 3-factor model prices portfolios sorted on characteristics, solving many of the anomalies with the CAPM. If stocks are priced rationally, systematic differences in average returns are due to differences in risk. The alphas that sorted portfolios have on the CAPM are due to accumulated risk premiums that do not come from exposure to the market portfolio. Since the inclusion of HML and SMB eliminates these alphas, these portfolios must proxy for exposure to aggregate risk factors of hedging concern to investors. This must be true, as otherwise they would not receive a systematic premium. Given their importance in asset pricing, there has been much debate over what aggregate risks the HML and SMB portfolios proxy for. Petkova (2006), Lettau and Ludvigson (2001), Vassalou (2003) and Liew and Vassalou (2000) relate HML and SMB to news about the future state of the economy and innovations to the investment opportunity set. But there is also a link between firm size, book-to-market equity and bankruptcy risk. Fama and French (1995) originally suggested that the persistently low profitability of high book-to-market and low size firms could be due to ‘distress’. Dichev (1998) finds that bankruptcy risk is negatively related to firm size and positively related to book-to-market equity, and in doing so stimulated much research into distress risk. See for example Campbell et al (2008), Griffin and Lemmon (2002), Vassalou and Xing (2004), Chava and PURMANANDAM (2010) and Kapadia (2011). Given the controversy in the debate on distress risk, and the importance of the HML and SMB portfolios to asset pricing theory, this chapter seeks to investigate the link between exposure to the HML and SMB portfolios and bankruptcy risk.

Petkova (2006) finds that innovations in the investment opportunity set are significantly correlated to HML and SMB, and that an ICAPM model that contains as factors the market return, and innovations to the aggregate dividend yield, term spread, default spread and 1-month T-bill yield, explains the cross-section of average returns on 25 portfolios sorted by firm size and book-to-market equity almost as well as the FF model. The market portfolio proxies for wealth, and it also proxies for a fraction of the changes to the investment opportunity set, but not enough to fully price portfolios of stocks sorted on characteristics. Stocks with extreme characteristics are relatively sensitive to changes in the investment opportunity set,
and require factors which strongly capture innovations thereof in order to be priced. Vassalou (2003) finds that news related to future GDP growth is an important factor for explaining the 25 portfolios. A model including this factor via a mimicking portfolio can explain returns about as well as the FF model. When the GDP news-related factor is included in the asset pricing model, SMB and HML lose most of their ability to explain returns, suggesting that much of the information in HML and SMB is news related to future GDP growth. Lettau and Ludvigson (2001) demonstrate that by allowing the parameters of the function relating the stochastic discount factor to the chosen factors to vary as a function of a conditioning variable reflecting investor expectations, $cay$, the CAPM, human capital CAPM, and consumption CAPM all achieve dramatically improved performance in terms of R-squared and significance of coefficients, when pricing the 25 portfolios. The most success is observed with the human capital CAPM and the consumption CAPM, where the R-squareds are comparably large to that achieved with the FF 3-factor model. Liew and Vassalou (2000) show that returns on HML and SMB are positively related to future growth in the real economy, and the predictive power of these portfolios is not diminished by the market portfolio when that is included. Inclusion of other business cycle variables does not eliminate the forecasting ability of HML and SMB either. The literature would seem to demonstrate that the HML and SMB factors are barometers of investor expectations regarding the future state of the economy and the investment opportunity set, and that covariates which proxy for this are what is required to price the 25 portfolios. Assuming rational expectations and efficient markets, this would suggest that value stocks and small stocks fundamentally perform relatively well when the economy is good, and relatively badly when the economy is bad.

But let us not forget the link between firm size, book-to-market equity and bankruptcy risk. Fama and French (1995) find that high book-to-market stocks have sustained low profitability for approximately five years before and after portfolio formation, while low book-to-market stocks have sustained high profitability. High book-to-market stocks have sustained high book-to-market ratios for approximately five years before and after portfolio formation, with the same pattern for low book-to-market. The evidence would appear to suggest that value stocks are discounted due to lower profitability, which gradually improves, leading to a reduction
in book-to-market ratio and their eventual departure from the value portfolio. Fama and French describe the persistent low profitability of the high book-to-market portfolio of firms as 'distress'. Dichev (1998) finds that bankruptcy risk (proxied by Z-score and O-score) is negatively related to firm size and positively related to book-to-market. Campbell et al (2008) use the fitted probability of failure from their logit model as a measure of financial distress, and create portfolios of stocks sorted by this probability to find that financially distressed firms have high market betas and high loadings on the HML and SMB factors.

Putting all the evidence together, we should therefore expect high betas on HML and/or SMB to be related to distress risk, as these will tend to be the small size and high value stocks that are most likely to be distressed and are fundamentally sensitive to the future state of the economy. This is a prediction that I investigate empirically in this paper.

I use a longitudinal dataset of firm-month observations to estimate a logit model for the probability of default, similar to Campbell et al (2008) and Shumway (2001). To avoid the use of noisy default risk proxies, I use actual incidences of default as the dependent variable. This requires that the analysis is restricted to the firm-level. I estimate Carhart (1997) 4-factor betas for every firm-month using 250-day windows of daily data, and use these as explanatory variables in the logistic regression. Since OLS coefficients are being used as regressors, naive logit coefficients are subject to bias analogous to the bias encountered in a 2-stage least squares regression. The SIMEX method is applied to correct for this measurement error bias. The results demonstrate that loadings on HML and SMB are significantly related to incidence of default. Surprisingly the sign of the coefficient on the HML beta is negative. To investigate further, nonlinear support vector machines (SVM) are applied to the same classification problem. One of the advantages of SVM is that the underlying equation need not be linear, as is the case for a logit specification, and moreover the SVM itself learns the optimal shape to use to separate the classes. These results clearly show that a high magnitude for a firm’s HML or SMB beta implies a higher probability of default. Stock returns are sensitive to the news proxied by the HML and SMB portfolios, either positively or negatively, when their fundamental financial performance is poor and they are at risk of bankruptcy.
Avramov et al (2013) explore the relationship between financial distress and asset pricing anomalies. In particular they examine financial distress as proxied by rating downgrades, because this has ex ante implications for a firm’s future performance. They find that the profitability of many trades exploiting anomalies is mostly due to the short side of the strategy among the lowest-rated firms. The value effect is only significant amongst the low-rated stocks, and corresponds to long positions that survive financial distress and go on to generate relatively high returns. As part of their study, they demonstrate that portfolios containing stocks with the highest (i.e. best) tercile of credit ratings have lower betas on MKT, SMB and HML, and that betas increase as the tercile of credit ratings worsens. This goes some way to addressing the prediction motivated earlier in this introduction. However although Avramov et al (2013, Table 1) demonstrate that $\beta_{HML}$ is hump-shaped in credit risk, the paper does not go further regarding this nonlinearity since its focus is on the profitability of trading strategies based on asset price anomalies. In contrast this study takes a closer look at the relationship between $\beta_{SMB}$, $\beta_{HML}$ and the ex ante probability of default as motivated by the prediction outlined earlier in this introduction, and furthermore elaborates on the nonlinear nature of that relationship.

Campbell et al (2008) estimate a default risk proxy using a logit specification, and sort stocks into portfolios according to that proxy before running 3-factor regressions on the portfolios. In Table VI and Figure 3 of that paper they demonstrate that $\beta_{HML}$ and to a lesser extent $\beta_{MKT}$ are hump-shaped in distress risk. Garlappi and Yan (2011) provide theoretical and empirical reasons to believe there is a nonlinear relationship between default probability and beta. They construct an equity valuation model involving financial leverage, wherein shareholders might be incentivized to default on their debt in order to recover part of the residual firm value following the resolution of financial distress. They then go on to use the distance-to-default measure as a proxy for a firm’s distress risk, and group stocks into portfolios according to that proxy to empirically examine how $\beta_{MKT}$ varies in distress risk. Similar to Campbell et al (2008) they find that beta is hump shaped as distress risk rises. In this study I estimate a logit that relates factor exposures to bankruptcy observations at the firm-level, exploiting the SIMEX method to overcome the measurement error issue. Similar to Griffin and Lemmon (2002) and
I use a logit specification with ex post bankruptcy observations as the left-hand side variable. As in those studies, the model predicted probabilities of default that result serve as a proxy for distress risk. I therefore examine the relationship between sensitivity to SMB and HML, and ex ante distress risk. The question that matters is whether distressed stocks are fundamentally sensitive to the future state of the economy. I then proceed to use support vector machines (SVM) with the aim of empirically verifying the relationship between Fama-French 3-factor loadings and distress risk, as measured by the estimated ex ante probability of default, while accounting for nonlinearity. As a result I demonstrate that stock returns are sensitive to aggregate macroeconomic risks when their fundamental financial performance is poor, and that this sensitivity can be either positive or negative.

II. The Basic Model

The study uses a limited dependent variable setting to regress a binary variable with values 1 for observed default and 0 otherwise, on lagged explanatory variables. Using the Carhart (1997) model, let’s denote the factor mimicking portfolios $R_{MKT}^{t}$, $R_{SMB}^{t}$, $R_{HML}^{t}$ and $R_{WML}^{t}$, corresponding to market, size, value and momentum factors. The factor loadings on $R_{MKT}^{t}$, $R_{SMB}^{t}$, $R_{HML}^{t}$ and $R_{WML}^{t}$ are denoted $eta_{MKT}^{i}$, $eta_{SMB}^{i}$, $eta_{HML}^{i}$ and $eta_{WML}^{i}$ respectively for each stock $i$, and are obtained from the following time-series regression:

$$R_{i}^{t} - R_{f}^{t} = \alpha^{i} + \beta_{MKT}^{i} (R_{MKT}^{t} - R_{f}^{t}) + \beta_{SMB}^{i} R_{SMB}^{t} + \beta_{HML}^{i} R_{HML}^{t} + \beta_{WML}^{i} R_{WML}^{t} + \epsilon_{i}^{t} (1)$$

Where $R_{i}^{t}$ is the return on stock $i$, and $R_{f}^{t}$ is the risk-free rate. The first step of the procedure calculates OLS estimates of these factor loadings for firm $i$ at time $t$ using a rolling window of the past 12 months of daily data. The aim of these rolling windows is to capture dynamic variation in a stock’s exposure to the factors as the firm’s level of distress changes. This is repeated for $i = 1, ..., N$ and $t = 1, ..., T$ to result in $T \times N$ individual observations of the vector $(\beta_{MKT}^{i,t}, \beta_{SMB}^{i,t}, \beta_{HML}^{i,t}, \beta_{WML}^{i,t})$, one vector for each firm-month. The $T \times N$ observations of
the vector \( \beta_{i,t}^{MKT}, \beta_{i,t}^{SMB}, \beta_{i,t}^{HML}, \beta_{i,t}^{WML} \) are then used as the explanatory variables in the linear relationship underlying the following logit model:

\[
Y_{i,t}^t = \begin{cases} 
1 & \text{if } -u_{i,t} < a + b\beta_{i,t}^{MKT} + c\beta_{i,t}^{SMB} + d\beta_{i,t}^{HML} + e\beta_{i,t}^{WML} \\
0 & \text{otherwise}
\end{cases}
\]

(2)

For \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \). \( Y_{i,t}^t \) is a binary variable taking value 1 if the firm defaults within the \( \tau \) months following time point \( t \), and 0 otherwise, where \( \tau \) is the lag chosen. In practice a range of lags from 1 month to 48 months could be used, but most studies, e.g. Campbell et al (2008) and Chava and Jarrow (2004), focus on using a 1-month ahead predictor for failure risk, and the same is done here. Since a lag of 1 month is chosen, it is natural for each \( t \) to also correspond to 1 month. Note that we are in a panel data setting, which will involve the maximum likelihood estimation considering \( T \times N \) individual observations of the input vector \( \left( \beta_{i,t}^{MKT}, \beta_{i,t}^{SMB}, \beta_{i,t}^{HML}, \beta_{i,t}^{WML} \right) \), i.e. \( T \times N \) firm-months, to estimate one parameter set \( (a, b, c, d, e) \). Where indicated in the results, Shumway (2001)'s market and accounting variables are used as controls for the logistic regression.

### III. Data

The bankruptcy indicator was taken from Sudheer Chava’s bankruptcy database\(^{10}\) which spans from 1962 to July 2008. It includes all bankruptcy filings in the Wall Street Journal Index, the SDC database, SEC filings, and the CCH Capital Changes Reporter. The stock market data was taken from CRSP, and covers all NYSE, AMEX and NASDAQ firms listed from 1961 till July 2008. Accounting variables were sourced from the CRSP-COMPUSTAT merged database. The 4-factor returns time series \( R^{MKT}, R^{SMB}, R^{HML} \) and \( R^{WML} \) were taken from Kenneth French’s online data library, as was the risk-free rate \( R^f \).

It is interesting to use a number of different versions of the stock sample. Firstly, Garlappi and Yan (2011) note that low price shares are often those in the greatest distress, and filtering

\(^{10}\)Chava and Jarrow (2004) and Chava et al (2011).
them out can decisively alter results (see Table I and Figure 3 from their paper). However on the other hand Chava and Purnanandam (2010) suggest that shares with market price < $1 suffer from market microstructure issues, such as bid-ask bounce and forced delisting, and consequently filter them out. Therefore I run one version of the experiment including low price shares, and another version excluding them. Secondly, Shumway’s model requires accounting data, but if you are only using the 4-factor betas and the total volatility as your explanatory variables then you do not require accounting data and can therefore include stocks without a COMPUSTAT match. Since using COMPUSTAT engenders a well-known survivorship bias, it is also interesting to see what happens with a dataset that does not require COMPUSTAT matches. Hence define four different versions of the full-sample:

- **D1**: CRSP NYSE/AMEX/NASDAQ firm-months with COMPUSTAT data and a market price $\geq$ $1$.
- **D2**: CRSP NYSE/AMEX/NASDAQ firm-months with COMPUSTAT data and a market price observation.
- **D3**: CRSP NYSE/AMEX/NASDAQ firm-months with a market price $\geq$ $1$.
- **D4**: CRSP NYSE/AMEX/NASDAQ firm-months with a market price observation.

Note that D1, D2, D3 and D4 have 505, 1093, 613 and 1253 defaults in their full-samples respectively, so another benefit from relaxing the requirements that each firm-month has a market price $> $1 and a COMPUSTAT match is to vastly increase the number of bankruptcies in the full-sample. Please see Table 9 for descriptive statistics of the bankruptcy data. Consistent with the distress risk literature, each explanatory variable is winsorized at the 1st and 99th percentiles of the distribution observed across all firm-months. This step is taken due to the presence of large outliers in accounting data. Please see Table 10 for descriptive statistics summarizing the winsorized explanatory variables. $NITA$ is the annual net income of a firm scaled by its total assets, $TLTA$ is the total liabilities of a firm scaled by its total assets, $EXRET$ is the return on a stock in excess of the return on a value-weighted CRSP

---

11 Albeit arguably only a minor bias e.g. see Chan et al (1995).
Table 9: Descriptive Statistics for Bankruptcy Dataset.

The bankruptcy data was taken from Chava and Jarrow (2004), and the interested reader is referred to that paper for additional summary statistics and analysis of the observed defaults. However note that here I use an updated version of the database that includes recent observations up to and including 2008. Firm-months with a refresh rate of less than 8.33% are excluded from these statistics. Please see the discussion of stale data later in this section for more details.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># firms in sample</td>
<td>25569</td>
</tr>
<tr>
<td># firms with no bankruptcy</td>
<td>24308</td>
</tr>
<tr>
<td># bankrupt firms</td>
<td>1261</td>
</tr>
<tr>
<td># bankrupt firms with price observation</td>
<td>1253</td>
</tr>
<tr>
<td># bankrupt firms with $p &lt; $1</td>
<td>640</td>
</tr>
<tr>
<td># firms with COMPUSTAT match</td>
<td>20974</td>
</tr>
<tr>
<td># bankrupt firms with COMPUSTAT match</td>
<td>1093</td>
</tr>
<tr>
<td># bankrupt with COMPUSTAT match &amp; price observation</td>
<td>1093</td>
</tr>
<tr>
<td># bankrupt with COMPUSTAT match &amp; $p &lt; $1</td>
<td>588</td>
</tr>
</tbody>
</table>

NYSE/AMEX/NASDAQ index, over the last month, and $RSIZE$ is the relative size, calculated as the log of the ratio of the firm’s market capitalization to the total market capitalization of all NYSE/AMEX/NASDAQ stocks on CRSP at that time. $SIGMA(R)$ is defined as the standard deviation of the last 60 market returns of the stock, calculated following the use of interpolation to handle missing price data. $SIGMA(e)$ is the standard deviation of the residuals from regression equation [1]. Similarly $\beta_{MKT}, \beta_{SMB}, \beta_{HML}$ and $\beta_{WML}$ are the betas from that regression. At this stage it is useful to define three important regressor combinations that appear at various times throughout this study:

- **E1**: $\beta_{MKT}, \beta_{SMB}, \beta_{HML}, \beta_{WML}, NITA, TLTA, EXRET, RSIZE$ and $SIGMA(e)$.
- **E2**: $\beta_{MKT}, \beta_{SMB}, \beta_{HML}, \beta_{WML}$ and $SIGMA(R)$.
- **E3**: $NITA, TLTA, EXRET, RSIZE$ and $SIGMA(e)$. 
E1 includes the 4-factor betas and Shumway’s variables as controls, and may be considered the full model. Since no two explanatory variables have a correlation greater than 0.5 in magnitude, it is safe to conclude that multicollinearity is not an issue. E2 does not include the controls and uses $SIGMA(R)$ instead of $SIGMA(e)$. The advantage of this stripped-down model is that it does not require accounting data. E3 is just Shumway (2001)’s market and accounting variables model.

It is worth briefly discussing the issue of stale data. In Campbell et al (2008), $SIGMA$ observations are calculated using a window of 3 months of daily return observations. They code an observation as missing if fewer than one twelfth of return observations are nonzero. In this study I follow suit. If a firm-month’s 250-day OLS window contains fewer than one twelfth nonzero daily return observations, then I exclude that firm-month from the study. Please see Figure 10 for a histogram that illustrates the refresh rates for the firm-months in the unfiltered full-sample, i.e. D4. The median refresh rate is 81%, and 2.81% of firm-months have a refresh rate of lower than 8.33%, corresponding to one twelfth. However in this study not only are we estimating stock volatilities, but we are also estimating 4-factor betas, and it is the latter that are more critical to the hypothesis tests. Even if we filter out firm-months with a refresh rate of lower than 8.33%, non-synchronous trading engenders the risk of understating betas, as in Epps (1979). Section VII looks at this issue more thoroughly.
Table 10: Descriptive Statistics for the Explanatory Variables.
Each explanatory variable is first winsorized at the 1st and 99th percentiles of the distribution observed across all firm-months. This step is taken due to the presence of large outliers in accounting data, and is consistent with the distress risk literature. $SIGMA(e)$ and $SIGMA(R)$ have been expressed at a monthly scale.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>St. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NITA</td>
<td>0.0068</td>
<td>-0.0084</td>
<td>-0.4142</td>
<td>0.1053</td>
<td>0.0704</td>
</tr>
<tr>
<td>TLTA</td>
<td>0.5349</td>
<td>0.5366</td>
<td>0.0311</td>
<td>1.2949</td>
<td>0.2663</td>
</tr>
<tr>
<td>RSIZE</td>
<td>-10.9834</td>
<td>-10.8784</td>
<td>-15.2740</td>
<td>-5.9172</td>
<td>2.0041</td>
</tr>
<tr>
<td>EXRET</td>
<td>-0.0084</td>
<td>-0.0020</td>
<td>-0.4004</td>
<td>0.5563</td>
<td>0.1407</td>
</tr>
<tr>
<td>$SIGMA(e)$</td>
<td>0.1110</td>
<td>0.1380</td>
<td>0.0001</td>
<td>0.5427</td>
<td>0.0984</td>
</tr>
<tr>
<td>$SIGMA(R)$</td>
<td>0.1135</td>
<td>0.1428</td>
<td>0</td>
<td>0.6107</td>
<td>0.1083</td>
</tr>
<tr>
<td>$\beta_{MKT}$</td>
<td>0.7992</td>
<td>0.8245</td>
<td>-0.7951</td>
<td>2.7185</td>
<td>0.6607</td>
</tr>
<tr>
<td>$\beta_{SMB}$</td>
<td>0.5736</td>
<td>0.6865</td>
<td>-1.2654</td>
<td>3.3515</td>
<td>0.8354</td>
</tr>
<tr>
<td>$\beta_{HML}$</td>
<td>0.1874</td>
<td>0.2018</td>
<td>-2.7205</td>
<td>3.0496</td>
<td>0.9280</td>
</tr>
<tr>
<td>$\beta_{WML}$</td>
<td>-0.0330</td>
<td>-0.0691</td>
<td>-2.4346</td>
<td>2.0237</td>
<td>0.6985</td>
</tr>
</tbody>
</table>

IV. Methodology

The methodology used in this study follows some of the key papers in the distress risk literature including Shumway (2001), Chava and Jarrow (2004) and Campbell et al (2008). In particular the statistical method behind the logit models follows Shumway (2001), wherein a panel data logit estimation using $T \times N$ firm-months is shown to be equivalent to a discrete-time hazard rate model, as required. A firm is included in the study from the first time at which all necessary market and accounting information are available, until either it experiences bankruptcy or is censored. Censoring occurs when market prices are no longer observed for the firm, e.g. because the firm has been acquired. If a company has multiple bankruptcy filings I consider only the first instance, and then remove them from the sample. Similar to Chava and Jarrow (2004),
missing observations of market or accounting data are handled by substituting the previous available observation. So for D1 and D2, if a particular firm-month is missing COMPUSTAT accounting data but that data was observed for that firm at a previous point in time, the firm-month is included. In the event that a stock is missing a lot of market price data, note that the worst-affected firm-months are excluded from the study altogether as a result of the rule on refresh rates discussed in section III. Consistent with the distress risk literature, each explanatory variable is winsorized at the 1st and 99th percentiles of the distribution observed across all firm-months. This is done separately for the in-samples, out-of-samples and full-samples. All explanatory data is lagged in the sense that when at time $t$ we are forecasting bankruptcy over the interval $(t, t + 1]$, only the most recent explanatory data available at or before time $t$ is used. Finally, since the factor loadings are estimated using OLS, standard theory indicates that they will suffer from estimation error. Referring to Carroll et al (2006) this will result in bias if $(\beta_{i,t}^{MKT}, \beta_{i,t}^{SMB}, \beta_{i,t}^{HML}, \beta_{i,t}^{WML})$ are then used as the explanatory variables for a logistic regression. Both the logit coefficients and their standard errors will be attenuated due to the bias. To mitigate this issue, one year’s worth of daily returns is used in the OLS regressions run using equation [1], so as to achieve as much convergence in the estimators as possible, while keeping the betas relevant as a predictor of default over the coming month. In this way it is hoped that the naive logit regressions will retain some validity, and I include the results below in section V since the method is widely known and easily understood. Nonetheless, a full correction for the measurement error issue is made in section VI using the SIMEX method.

V. Logit Results
Table 11: Naive Logit Regressions for Default.

A logit regression is used to relate the panel data of default indicators to the winsorized explanatory variables. This was done using different versions of the sample, denoted D1, D2, D3 and D4, and for different regressor combinations as indicated. Absolute values of the t-stats are shown underneath in parentheses.

ROC denotes the Receiver Operating Characteristic. This measures the out-of-sample forecasting accuracy of the model, and a score of 1 corresponds to perfect prediction capability, while a score of 0.5 suggests the model is no better than randomly guessing. Please see the text for more details. $R^2$ denotes McFadden’s pseudo-$R^2$. 

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D1</th>
<th>D2</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-14.14(34.51)</td>
<td>-10.52(92.81)</td>
<td>-14.78(48.81)</td>
<td>-9.86(131.55)</td>
<td>-10.68(104.30)</td>
<td>-10.00(144.15)</td>
</tr>
<tr>
<td>$NITA$</td>
<td>-4.39(10.04)</td>
<td>-1.88(7.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TLTA$</td>
<td>3.80(20.20)</td>
<td>3.28(28.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RSIZE$</td>
<td>-0.15(4.23)</td>
<td>-0.29(11.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$EXRET$</td>
<td>-3.93(13.40)</td>
<td>-3.44(18.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SIGMA(e)$</td>
<td>1.51(10.95)</td>
<td>1.08(14.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SIGMA(R)$</td>
<td>2.64(30.92)</td>
<td>2.28(52.62)</td>
<td>2.85(35.56)</td>
<td>2.44(57.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{MKT}$</td>
<td>0.25(3.56)</td>
<td>0.14(2.17)</td>
<td>0.18(4.37)</td>
<td>0.04(1.08)</td>
<td>0.16(2.61)</td>
<td>0.05(1.28)</td>
</tr>
<tr>
<td>$\beta_{SMB}$</td>
<td>-0.02(0.38)</td>
<td>-0.05(1.04)</td>
<td>0.06(2.15)</td>
<td>0.03(1.19)</td>
<td>-0.02(0.46)</td>
<td>0.04(1.45)</td>
</tr>
<tr>
<td>$\beta_{HML}$</td>
<td>-0.01(0.16)</td>
<td>0.05(1.43)</td>
<td>-0.05(2.04)</td>
<td>0.01(0.64)</td>
<td>0.03(0.82)</td>
<td>0.00(0.10)</td>
</tr>
<tr>
<td>$\beta_{WML}$</td>
<td>-0.21(4.36)</td>
<td>-0.25(5.31)</td>
<td>-0.18(6.50)</td>
<td>-0.21(7.56)</td>
<td>-0.21(4.90)</td>
<td>-0.20(7.55)</td>
</tr>
<tr>
<td>ROC</td>
<td>0.78(5.31)</td>
<td>0.79(6.50)</td>
<td>0.88(7.56)</td>
<td>0.89(4.90)</td>
<td>0.81(7.55)</td>
<td>0.90(5.31)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.17(1.43)</td>
<td>0.10(2.04)</td>
<td>0.22(2.64)</td>
<td>0.15(3.02)</td>
<td>0.11(3.56)</td>
<td>0.15(2.17)</td>
</tr>
</tbody>
</table>

67
Table 11 reports the logit coefficients along with absolute values of their t-stats in parentheses underneath. The $R^2$ reported is McFadden’s pseudo-$R^2$, calculated as $1 - \frac{L_1}{L_0}$, where $L_1$ is the log likelihood of the estimated model and $L_0$ is the log likelihood of a null model that includes only a constant term. All of these statistics were estimated using the full-sample of data. The $ROC$ is the Receiver Operating Characteristic. To calculate this, first the full-sample is split into an in-sample (IS) and an out-of-sample (OOS), where the cut-off point between them has been chosen as January 2000. The model is then trained using the IS to estimate coefficients. This IS model is used to estimate a probability of default over the forthcoming 1-month period for each firm-month in the OOS. Each firm-month is then ranked from highest probability of default to lowest, and the sum of the actual defaults is taken so as to obtain the cumulative % of actual defaults as a function of the probability of default (highest to lowest). The ROC is the integral of the cumulative % of actual defaults as a function of the probability of default. Therefore an $ROC$ of 1 corresponds to perfect OOS prediction capability, while 0.5 would suggest that the model is no better than randomly guessing. This version of the $ROC$ is sometimes referred to as a Cumulative Accuracy Profile (CAP).

It is notable that when low price shares are included (D2) and the full specification is used including all control variables (E1), the 4-factors are statistically significant predictors of default over a 1-month horizon. It might be surprising to see that probability of default is decreasing in $\beta_{HML}$, and I take a closer look at why this might be in section VIII. It would appear that the inclusion of shares with market price < $1 is decisive to the significance of $\beta_{SMB}$ and $\beta_{HML}$, at least for the case of a logit specification with a linear underlying equation. Referring to Table 9, of those 1093 bankruptcies with a COMPUSTAT match, 588 have $p < $1 at the time of bankruptcy. Therefore excluding low price shares removes more than half of the bankruptcy observations. This reduction in sample size might be the reason for the loss of significance of $\beta_{SMB}$ and $\beta_{HML}$ in D1. However it is also quite possible that the relationship identified is specific to low price shares. The implication would be that low price shares that are at risk of bankruptcy have a greater sensitivity to the news proxied by SMB and HML than higher price shares that are also at risk of bankruptcy. In section VIII I explore whether a nonlinear model is robust to the exclusion of low price shares.
Table 12 investigates what happens if we try to predict bankruptcies further into the future. Using the full specification (E1) and including low price shares (D2) I estimate the probability of bankruptcy in 1 month, 6 months, 1 year, 2 years and 4 years. Note that similarly to Campbell et al (2008), at time $t$ I am estimating the probability of bankruptcy occurring over the interval $(t + j - 1, t + j]$, conditional on no default occurring over the interval $(t, t + j - 1]$, where $j \in [1, 6, 12, 24, 48]$. Table 12 confirms that the significance of the coefficients is robust to longer horizons.
Table 12: Naive Logit Regressions for Default at Longer Horizons.

A logit regression is used to relate the panel data of default indicators to the winsorized explanatory variables. This was done using the full-sample that includes low price shares (D2), and the full specification including all control variables (E1). The model predicts the probability of bankruptcy occurring over the interval \((t + j - 1, t + j]\), where \(j\) is as indicated. Absolute values of the \(t\)-stats are shown underneath in parentheses. ROC denotes the Receiver Operating Characteristic. This measures the out-of-sample forecasting accuracy of the model, and a score of 1 corresponds to perfect prediction capability, while a score of 0.5 suggests the model is no better than randomly guessing. Please see the text for more details. \(R^2\) denotes McFadden’s pseudo-\(R^2\).

<table>
<thead>
<tr>
<th>(j) (months)</th>
<th>1</th>
<th>6</th>
<th>12</th>
<th>24</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(48.81)</td>
<td>(55.48)</td>
<td>(55.26)</td>
<td>(49.84)</td>
<td>(42.31)</td>
</tr>
<tr>
<td>(NITA)</td>
<td>-1.88</td>
<td>-2.06</td>
<td>1.33</td>
<td>-0.87</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(7.90)</td>
<td>(9.30)</td>
<td>(5.49)</td>
<td>(3.01)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>(TLTA)</td>
<td>3.28</td>
<td>2.91</td>
<td>2.39</td>
<td>1.72</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>(28.15)</td>
<td>(29.57)</td>
<td>(25.58)</td>
<td>(17.80)</td>
<td>(10.51)</td>
</tr>
<tr>
<td>(RSIZE)</td>
<td>-0.29</td>
<td>-0.24</td>
<td>-0.20</td>
<td>-0.13</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(11.79)</td>
<td>(10.67)</td>
<td>(7.05)</td>
<td>(3.80)</td>
<td></td>
</tr>
<tr>
<td>(EXRET)</td>
<td>-3.44</td>
<td>-1.52</td>
<td>-1.02</td>
<td>-0.84</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>(18.97)</td>
<td>(10.80)</td>
<td>(7.37)</td>
<td>(5.55)</td>
<td>(1.17)</td>
</tr>
<tr>
<td>(SIGMA(e))</td>
<td>1.08</td>
<td>0.86</td>
<td>0.73</td>
<td>0.57</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(14.29)</td>
<td>(12.78)</td>
<td>(10.63)</td>
<td>(7.43)</td>
<td>(4.96)</td>
</tr>
<tr>
<td>(\beta_{MKT})</td>
<td>0.18</td>
<td>0.24</td>
<td>0.26</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(4.37)</td>
<td>(6.23)</td>
<td>(6.44)</td>
<td>(6.67)</td>
<td>(5.47)</td>
</tr>
<tr>
<td>(\beta_{SMB})</td>
<td>0.06</td>
<td>0.07</td>
<td>0.15</td>
<td>0.18</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(2.63)</td>
<td>(5.50)</td>
<td>(5.60)</td>
<td>(3.77)</td>
</tr>
<tr>
<td>(\beta_{HML})</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(2.04)</td>
<td>(1.67)</td>
<td>(1.92)</td>
<td>(4.18)</td>
<td>(2.11)</td>
</tr>
<tr>
<td>(\beta_{WML})</td>
<td>-0.18</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(6.50)</td>
<td>(6.10)</td>
<td>(4.47)</td>
<td>(2.47)</td>
<td>(2.02)</td>
</tr>
<tr>
<td>ROC</td>
<td>0.88</td>
<td>0.84</td>
<td>0.80</td>
<td>0.72</td>
<td>0.65</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.22</td>
<td>0.14</td>
<td>0.09</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Although the results of the naive logits are promising, please read on to Sections VI and VIII for some more conclusive evidence. However it is worth taking the time to briefly mention one consequence of the naive logit tests. Referring to Table 13, the results show that the 4-factor model used in conjunction with the total volatility (E2) can be used to achieve marginally better OOS forecasting accuracy than Shumway’s market and accounting variable model (E3), with the added advantages that it only requires market variables and is therefore applicable to a much wider range of stocks, and can produce revised forecasts at a much higher frequency.

In unreported results I show that it is \( SIGMA(R) \) and \( \beta_{WML} \) which provide almost all of the explanatory power in the 4-factor model. It would seem that stocks with negative momentum and high total volatility are most at risk of default, something which is borne out in results from Shumway (2001) and Chava and Jarrow (2004).

Table 13: Default Prediction Using the 4-factor Betas and \( SIGMA(R) \).

Logit regressions were used to relate the panel data of default indicators to the winsorized explanatory variables. This was done using a specification matching Shumway (2001)’s market and accounting variable model (E3), and a specification that uses only the 4-factor betas and the total volatility (E2). This was repeated using a version of the dataset that excludes low price shares (D1), and a version that includes low price shares (D2). \( ROC \) denotes the Receiver Operating Characteristic. This measures the out-of-sample forecasting accuracy of the model, and a score of 1 corresponds to perfect prediction capability, while a score of 0.5 suggests the model is no better than randomly guessing. Please see the text for more details. \( R^2 \) denotes McFadden’s pseudo-\( R^2 \).

<table>
<thead>
<tr>
<th></th>
<th>D1E3</th>
<th>D1E2</th>
<th>D2E3</th>
<th>D2E2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC</td>
<td>0.7796</td>
<td>0.7927</td>
<td>0.8795</td>
<td>0.8900</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.1705</td>
<td>0.1038</td>
<td>0.2159</td>
<td>0.1457</td>
</tr>
</tbody>
</table>

VI. Simulated Extrapolation
The naive logit regressions from Section V provide an heuristic insight into the relationship between probability of default and betas, but since \( (\beta_{i,t}^{\text{MKT}}, \beta_{i,t}^{\text{SMB}}, \beta_{i,t}^{\text{HML}}, \beta_{i,t}^{\text{WML}}) \) were themselves estimated in OLS regressions, they are effectively measured with error. This will lead to attenuated logit coefficients, and understated standard errors. Please see Carroll et al (2006) and Wang et al (1998) for more details on measurement error in logistic regression.

Wang et al (1999) explain how regression calibration can be used to correct for the coefficient attenuation resulting from measurement error bias in logistic regression. This approach is advantageous since it is not computationally intensive. Unfortunately the asymptotic limits are obtained for a case where the number of clusters approaches infinity for fixed and finite subjects, i.e. \( T \gg N \). In this study the opposite is true, and the number of subjects far exceeds the number of clusters, i.e. \( N \gg T \). The next best alternative is known as Simulated Extrapolation (SIMEX). Wang et al (1998) provide a clear explanation of how to apply it to a longitudinal logistic regression where the number of subjects exceeds the number of clusters, i.e. \( N \gg T \). To briefly outline the procedure, let us denote \( X \) the true, unobserved covariate, and \( W \) the related error-prone covariate. In an additive error setting:

\[
W_{i,j} = X_{i,j} + U_{i,j}
\]

The \( U_{i,j} \) are independent of the \( X_{i,j} \) and are independently distributed \( N(0, \Sigma_{i,j}) \). We then proceed by estimating \( \Sigma_{i,j} \), which in this case comes directly from the OLS theory. Note that the technique can handle heteroskedasticity, as every firm-month \((i, j)\) can have its own covariance matrix \( \Sigma_{i,j} \). The procedure then calls for us to choose some positive value \( \xi \), and proceed to remeasure \( W \), \( B \) times as follows:

\[
W_{b,i,j} = W_{i,j} + \sqrt{\xi} U_{b,i,j}
\]

Note that:

\[
E [W_{b,i,j}(\xi)|X_{i,j}] = X_{i,j}
\]
And:

$$\text{var} [W_{b,i,j}(\xi)|X_{i,j}] = (1 + \xi)\Sigma_{i,j} = (1 + \xi)\text{var} [W_{i,j}|X_{i,j}]$$

Formally, the key property of the remeasured data is as follows:

$$\text{MSE} [W_{b,i,j}(\xi)] = E \left[ (W_{b,i,j}(\xi) - X_{i,j})^2 | X_{i,j} \right] \to 0 \text{ as } \xi \to -1$$

The logit specification is then estimated $B$ times, using the randomly generated $W_{b,i,j}$. This results in $B$ sets of logit coefficients, and the average is taken of each coefficient to arrive at the logit coefficient set corresponding to the $\xi$ value chosen: $\hat{\Theta}(\xi)$. This is repeated for a range of $\xi$ values, so that a function relating the logit coefficient values to $\xi$ can be estimated. Therefore by extrapolating the $\hat{\Theta}(\xi)$ function back to $\xi = -1$ we obtain a value of the coefficient for the hypothetical scenario where there is no measurement error, $\hat{\Theta}_{\text{simex}}$. This is the value of $\xi$ for which $(1 + \xi)\Sigma_{i,j} = 0$ for all $i$ and all $j$. The SIMEX method naturally extends to estimation of measurement error-corrected standard errors. For more details the interested reader is referred to Carroll et al (2006), Chapter 5 and Appendix B.4.

One drawback of the SIMEX method is that it can rapidly become computationally very intensive. Therefore the $B$ and $\xi$ parameter choices were made so as to compromise between computational feasibility and the efficacy of the SIMEX method. Setting $B = 200$ and choosing six $\xi$ values, $(0, 0.4, 0.8, 1.2, 1.6, 2)$, requires 1001 logit models to be run, i.e. 200 for each non-zero $\xi$ and one for $\xi = 0$. Since each dataset version contains approximately three million firm-months, the procedure also requires the generation of approximately three billion sets of four correlated random variables. All of this is computationally very intensive, but referring to Carroll et al (2006) such parameter choices are quite typical. Consequently the procedure is only applied to the full specification (E1) using the dataset including low price shares (D2). Furthermore $\text{ROC}$ are not reported, due to the computation time that would be involved in estimating a set of in-sample coefficients. Nonetheless please note that the $\text{ROC}$ for the naive
logit models in Table 11 already validate the predicting model used. Please see Table 14 for the SIMEX results.

Table 14: SIMEX Logit Regression for Default.
A SIMEX logit regression was used to relate the panel data of default indicators to the winsorized explanatory variables. Due to the computational cost of the SIMEX method, this was done only once, using the full-sample including low price shares (D2), and the full specification using all control variables (E1). Absolute values of the $t$-stats are shown underneath in parentheses. $R^2$ denotes McFadden’s pseudo-$R^2$.

<table>
<thead>
<tr>
<th>Intercept</th>
<th>NITA</th>
<th>TLTA</th>
<th>RSIZE</th>
<th>EXRET</th>
<th>SIGMA(e)</th>
<th>$\beta_{MKT}$</th>
<th>$\beta_{SMB}$</th>
<th>$\beta_{HML}$</th>
<th>$\beta_{WML}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15.24</td>
<td>-1.81</td>
<td>3.28</td>
<td>-0.32</td>
<td>-3.35</td>
<td>0.98</td>
<td>0.29</td>
<td>0.10</td>
<td>-0.08</td>
<td>-0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>(46.58)</td>
<td>(7.56)</td>
<td>(28.10)</td>
<td>(12.08)</td>
<td>(18.51)</td>
<td>(12.41)</td>
<td>(4.85)</td>
<td>(2.38)</td>
<td>(2.24)</td>
<td>(7.54)</td>
<td></td>
</tr>
</tbody>
</table>

Comparing Table 14 to Table 11, there are no meaningful differences in the sign, magnitude or significance of the logit coefficients, and hence the conclusions from Table 11 are robust to corrections for the measurement error issue.

VII. Non-Synchronous Data

So far all the tests have involved betas that were estimated using daily return observations, the aim being to mitigate measurement error issues. However this is not without its drawbacks. The issue is that increasing the sampling frequency for returns leads to a bias towards zero in covariation statistics. See Epps (1979) for example. For highly traded stocks there is not much of an issue when using daily data, but the betas of small stocks, with less frequent trading, can be substantially underestimated when using daily returns. Barndorff-Nielsen et al (2008) provide a survey of the literature on non-synchronous data issues.

For this reason the experiment was repeated using weekly return observations to estimate the factor loadings. I used the full specification (E1) and the dataset that includes low price shares (D2) to estimate a naive logit and a SIMEX-corrected logit. Referring to Table 15, it would appear that the significance of the logit coefficients does not survive when weekly returns are used for the beta estimations. This appears to be the case even with the SIMEX correction.
and raises the possibility that the results from Table 11 could be sensitive to non-synchronous data as well as low price shares.

However note that a logit model uses a linear underlying equation and will therefore only successfully separate bankruptcies from non-bankruptcies if there is a monotonically increasing or decreasing relationship between the probability of bankruptcy and the regressor. In this case however there is existing evidence that gives us reason to believe that high magnitudes of $\beta_{HML}$ might be related to increased probability of bankruptcy. For example Griffin and Lemmon (2002, Table 1) show that the quintile of stocks with the highest probability of default also contains the greatest range of book-to-market values. Any linear model is therefore potentially misspecified. Please read on to Section VIII, where I examine the evidence using a nonlinear model. Using SVM to apply a nonlinear underlying equation I demonstrate a relationship between the factor loadings and bankruptcy risk that is robust to non-synchronous data and the inclusion of low price shares.

Table 15: Logit Regressions for Default Using Weekly Return Observations: Non-Synchronous Data.

A naive logit regression and a SIMEX logit regression were used to relate the panel data of default indicators to the winsorized explanatory variables. Here, the 4-factor betas and the total volatility were estimated using weekly rather than daily returns, as a way to control for non-synchronous data. Due to the computational cost of the SIMEX method, this was done only using the full-sample including low price shares (D2), and the full specification using all control variables (E1). Absolute values of the t-stats are shown underneath in parentheses. ROC denotes the Receiver Operating Characteristic. This measures the out-of-sample forecasting accuracy of the model, and a score of 1 corresponds to perfect prediction capability, while a score of 0.5 suggests the model is no better than randomly guessing. Please see the text for more details. $R^2$ denotes McFadden’s pseudo-$R^2$.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>NITA</th>
<th>TLTA</th>
<th>RSIZE</th>
<th>EXRET</th>
<th>SIGMA</th>
<th>$\beta_{MKT}$</th>
<th>$\beta_{SMB}$</th>
<th>$\beta_{HML}$</th>
<th>$\beta_{WML}$</th>
<th>ROC</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td>-14.86</td>
<td>-1.74</td>
<td>3.29</td>
<td>-0.30</td>
<td>-3.46</td>
<td>1.32</td>
<td>0.03</td>
<td>-0.08</td>
<td>-0.12</td>
<td>0.88</td>
<td>0.88</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(52.76)</td>
<td>(7.16)</td>
<td>(28.04)</td>
<td>(13.07)</td>
<td>(18.92)</td>
<td>(14.42)</td>
<td>(1.14)</td>
<td>(4.76)</td>
<td>(0.95)</td>
<td>(6.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMEX</td>
<td>-15.03</td>
<td>-1.70</td>
<td>3.29</td>
<td>-0.32</td>
<td>-3.42</td>
<td>1.20</td>
<td>0.05</td>
<td>0.13</td>
<td>-0.02</td>
<td>0.22</td>
<td>NA</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(51.11)</td>
<td>(6.98)</td>
<td>(27.91)</td>
<td>(13.19)</td>
<td>(18.62)</td>
<td>(12.62)</td>
<td>(1.24)</td>
<td>(5.17)</td>
<td>(0.90)</td>
<td>(7.72)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VIII. Nonlinear Support Vector Machines
As mentioned previously Garlappi and Yan (2011) provide theoretical and empirical reasons to believe there is a nonlinear relationship between default probability and beta. In Table II of their paper they present empirical results showing CAPM betas are hump shaped as distress risk increases, and Figure 3 shows empirical evidence that the value spread is non-monotonic as a function of distress decile. It is unfortunate then that typical binary classifiers are only able to achieve a linear separation between classes. Although the probability gradient between class 0 and class 1 is nonlinear due to the cumulative density function used, the separation in input space is forced to be linear. In other words the underlying equation in [2] is linear. Obviously among the advantages of using logit and probit models are that they offer widely understood statistical tests for significance, and estimated coefficients that are easily interpreted as marginal effects. But when applied to data where there is known theoretical and empirical evidence of nonlinearity, those benefits may be outweighed by the risk that in fitting a linear separator to a nonlinearly separated dataset, spurious results may arise.

As a first step in addressing this reasonable concern, I split the full-sample of firm-months into those where a default occurs over the next month and those where no default occurs over the next month. Figures 11-16 show histograms that compare the distributions of the winsorized betas for the default firm-months vs. the non-default firm-months. These graphs indicate that the kurtosis of the beta distribution is much higher for those firm-months experiencing imminent default. It might be surprising to observe that it is not just high betas that are linked to default, but extreme magnitudes of betas. This nonlinear relationship was not apparent from the logit models, and in particular note that large positive $\beta_{HML}$ are linked to default, as are large positive $\beta_{SMB}$ and $\beta_{MKT}$. In order to handle this nonlinearity one option is to apply a nonlinear support vector machine. One of the advantages of SVM is that the procedure will separate the classes via an algorithm that chooses the shape of the separating boundary in an entirely data-dependent way. Another alternative would be to include regressors such as $\beta_{HML}^2$ in a logit model as a way to capture the nonlinearity, but the drawback is that this entails assuming a particular shape for the decision boundary. Since specification E1 involves nine regressors, we are therefore in 10-dimensional space and it is not possible to know what the shape of the nonlinear separating surface is. For this reason the SVM method
was favoured. The trade-off is that while the logit method provides tests for the statistical significance of coefficients, the SVM does not.

Alpaydin (2010) provides a good introduction to nonlinear support vector machines, while Suykens et al (2002) provides the necessary detail for what follows in this chapter. A Gaussian kernel was used. The two relevant parameters for the optimization (cost, $C$, and sigma, $\sigma$) were chosen using a 3-fold cross-validation applied to the in-sample (IS). More folds would have been used ideally, but the process is very computationally intensive given the size of the dataset. Not only are there approximately 2 million vectors in the in-sample, but since the number of defaults is far lower than the number of non-defaults, upsampling must be used. Upsampling ensures that the machine achieves roughly the same in-sample error rate for both the minority and majority class, by resampling the minority class until the proportions are equivalent. It is a technique commonly used in machine learning to ensure that the majority class does not overwhelm the optimization process, and to avoid a result with a very high success rate for the majority class, but a very low success rate for the minority class. An upsampled dataset will achieve a better $ROC$ statistic than a dataset with a disproportionate minority class. Following upsampling the in-sample alone contains as many as 3 million observation vectors, and hence a large-scale algorithm implementing the Nystrom method is required. The interested reader is referred to Suykens et al (2002) for more details of this technique. In a nutshell, the standard (small-scale) nonlinear support vector machine (SVM) algorithm optimizes a cost function that involves a $(T \times N) \times (T \times N)$ matrix of the kernel interactions between each firm-month vector in the in-sample and every other vector in the in-sample, known as the Gram matrix. If there are 3 million vectors in the in-sample, the associated Gram matrix has 9 trillion elements, which ceases to be computationally feasible. The Nystrom method provides a feasible and effective alternative.

To verify the robustness of an SVM model\textsuperscript{12}, a receiver operating characteristic ($ROC$) can be calculated to quantify the OOS prediction accuracy achieved by the model after being trained on the IS. This statistic is calculated in a similar way as for the logit models before. A logistic function is applied to the output of the SVM discriminant function so that rather

\textsuperscript{12}The choice of kernel function and explanatory variables, as applied to the dataset in question.
than the SVM making discrete binary predictions, we instead obtain a predicted probability of default. This procedure is known as Platt (2000) scaling. Following the OOS accuracy test, the SVM is finally trained on the full-sample (FS) to exploit all the data. One drawback with SVM vs. logit models is that it is not possible to describe the predictions of the model using a table of coefficients. The prediction for any instance of a vector that you may wish to classify is the result of a nonlinear function of the kernels between the instance you wish to classify, and a relevant subset of the vectors from the in-sample. Instead, one can graph how the predicted probability of default varies as we vary one explanatory variable of interest, while holding the others fixed. The issue of where to hold the other explanatory variables fixed is important, since the whole point of the exercise is that the decision boundary is nonlinear. I hold all explanatory variables fixed at their post-winsorized medians, and then vary the target variable from its post-winsorized minimum to maximum. Please see Figures 17-25 at the end of this document.

Strong ROC statistics suggest that the SVM models used are robust and have not been overfitted in-sample. For the full specification (E1) we see that increasing $\beta_{MKT}$, $\beta_{SMB}$ or $\beta_{HML}$ is linked to increasing probability of default. This result is robust to whether or not low price shares are included, and to whether daily or weekly returns are used to estimate betas. Therefore in this nonlinear setting we have a relationship between factor loadings and probability of firm default that is robust to the exclusion of low price shares, and is robust to non-synchronous data issues. One caveat to these results is that each graph is merely a 2D-slice through 10-dimensional space. Any of the relationships depicted in the graphs may change as the fixed variables are moved away from their medians. Nonetheless holding the variables fixed at their medians seems reasonable as a way to explore how the variation in the target variable affects probability of default for typical values of the other variables.

Unfortunately the SVM methodology does not provide tests for the statistical significance of individual regressors in the same way that logit models do. Instead the procedure relies on judicious choice of regressors followed by out-of-sample testing. In this case the choice of regressors was justified using the histograms in Figures 11-16, and the ROC then verifies how
robust the model is out-of-sample. The economic significance of a regressor is observed by the change in predicted probability of default given a change in the regressor. Similar to logit models, the marginal impact on predicted probability of changing a regressor is not constant. What we do observe is that holding all other regressors fixed at their winsorized medians, and varying either $\beta_{SMB}$ or $\beta_{HML}$ from their respective minimum observed value through to their maximum, results in predicted probabilities of default that vary over at least a 4% range in all cases. For example in Figure 17 the predicted probability of default varies from 34.73% to 38.83%. That is a wide variation in predicted probability of default, given that referring to Table 9 the unconditional probability of observing a default in the sample of firms with a COMPUSTAT match is 5.21%. The prediction capability of the models has of course been verified by ROC statistics.

The results suggest that loadings on the size and value factors are related to default. However, as might be expected given the controversy in the literature, the picture is more complicated than this. Robust to all versions we see a concave up parabola for likelihood of default as a function of $\beta_{HML}$ and $\beta_{SMB}$. So as the $\beta_{HML}$ or $\beta_{SMB}$ passes the minimum level of distress probability (usually at betas of -1 to 1), and becomes increasingly negative, the likelihood of default also increases. High magnitudes for a firm’s $\beta_{HML}$ or $\beta_{SMB}$ are related to increased probability of default. This might explain why the logit coefficients for $\beta_{HML}$ and $\beta_{SMB}$ are not robust for all versions of the study. Firms that are expected to receive higher returns than average via the factor loadings are linked to increased probability of default, and firms that are expected to receive lower returns than average via the factor loadings are also linked to default.

IX. Conclusions

This chapter examines the relationship between exposure to the HML and SMB portfolios and probability of default. I regress actual incidences of default on 4-factor betas using Shumway (2001)’s variables as controls. The results clearly show that high magnitudes for a firm’s $\beta_{HML}$ or $\beta_{SMB}$ are related to probability of default. What may be surprising is that a large negative
coefficient with HML and SMB also leads to rising default probability. Petkova (2006), Lettau and Ludvigson (2001), Vassalou (2003) and Liew and Vassalou (2000) find that the SMB and HML portfolios are sensitive to news about the future state of the economy and innovations to the investment opportunity set. In this chapter I demonstrate that stock returns are sensitive to these aggregate macroeconomic risks when their fundamental financial performance is poor, and that this sensitivity can be either positive or negative.
Chapter 4
Low Risk Investing

A long-short portfolio that buys low beta stocks and shorts high beta stocks has been shown to earn an anomalous premium on standard factor models. This chapter finds that for the 1990-2012 subsample the anomaly disappears when an unconditional 4-factor model is used, while for both the 1926-2012 full-sample and the 1926-1990 subsample the anomaly survives. The structural break could be attributable to a rise in leverage, as in Frazzini and Pedersen (2013). Furthermore, a conditional 4-factor model results in substantially reduced alphas for the full-sample and the 1990-2012 subsample. Conditional estimates are obtained using daily data and Lewellen and Nagel (2006)’s short-window regressions.
I. Introduction

Frazzini and Pedersen (2013, henceforth FP) show that for a sample of US equities 1926-2012, the risk-adjusted returns on portfolios of stocks sorted according to their CAPM beta is monotonically decreasing as the portfolio beta increases. Baker, Bradley and Wurgler (2010, 2013) show that the geometric mean return on portfolios sorted into quintiles according to their CAPM betas is monotonically decreasing as the portfolio beta increases. These findings are at odds with the financial axiom that investors are rewarded with higher returns for bearing systematic risk.

FP explain their results via a theoretical model that builds on the earlier work of Black (1972). Both of these papers relax the CAPM’s traditional assumption that investors are free to borrow without constraint. Black explores a model where there is a riskless asset available, but in which investors are not allowed to take short positions in the riskless asset. The result is a security market line (SML) that is shallower in slope than for the traditional CAPM. FP advance on this to suggest that the majority of investors, such as individuals, pension funds and mutual funds, are constrained in their borrowing and choose instead to overweight risky securities to increase returns. The result of this demand for high-beta stocks is that in equilibrium these stocks require lower risk-adjusted returns than low-beta stocks, as the latter are only appealing to those investors who can exploit leverage to amplify the returns. One of the predictions of their model is a flatter SML as in Black (1972). FP then go on to demonstrate that a market-neutral long-short portfolio going long on low beta stocks and short on high beta stocks (the “Betting-Against-Beta” factor) earns a statistically significant premium on standard factor models. Furthermore they provide evidence consistent with their leverage explanation.

An alternative explanation is given by Hong and Sraer (2012), who use differences-in-beliefs to explain the anomaly as the result of speculative overpricing of high beta stocks. When investors disagree about the common factor in firm cash-flows, high beta stocks experience a greater divergence of opinion about their payoffs, and short-sales constraints result in the high beta stocks being overpriced.
There is a literature that seeks to explain older and more familiar anomalies such as the value effect using a conditional CAPM. Jagannathan and Wang (1996) demonstrate that if a conditional CAPM is the true process generating returns, estimating coefficients using an unconditional regression will result in a biased alpha. The bias arises due to the covariance between the conditional risk premium and the conditional beta, a phenomenon referred to in this paper as "dynamic exposure". Unfortunately estimating conditional factor models is not without its difficulties. Ang and Chen (2007) find that a conditional one-factor model can explain the returns on stocks sorted according to book-to-market over their extended sample of 1926-2001, implying that it is dynamic exposure to the market premium which accounts for the unconditional alpha observed on value portfolios. However their study involves specifying the conditional beta as an endogenous latent process. Petkova and Zhang (2005) find that a portion of the premium on the HML portfolio is due to dynamic exposure to the market risk premium, but their study relies on the assumption that instruments such as the aggregate dividend yield, default spread, term spread and short-term interest rate are sufficient to capture investors' information sets. Lewellen and Nagel (2006, henceforth LN) provide a feasible method for estimating conditional factor models that circumvents these issues, and this is key to the results of this chapter.

This chapter seeks to address whether a conditional 4-factor model is capable of explaining the returns on beta-sorted long-short portfolios. I use CRSP data from 1926-2012 to run a simple version of the Betting-Against-Beta strategy that goes long on low beta stocks and short on high beta stocks. The returns on these portfolios are regressed on Carhart (1997)'s 4-factor model using standard OLS to obtain unconditional alphas. I then estimate conditional 4-factor alphas using LN's short-window regressions, and find that the conditional alphas are substantially smaller than the unconditional alphas. Following this I form subsamples, and find that for the 1990-2012 subsample the unconditional alphas are statistically insignificant. Furthermore, for this subsample the conditional alphas of the portfolios are orders of magnitude smaller that the unconditional alphas.

The evidence is that following 1990 the Betting-Against-Beta premium disappears. The structural break is coincident with economically significant growth in hedge fund AUM, which
is strongly suggestive of a rise in the proportion of capital under the control of leverage unconstrained arbitrageurs as defined by FP. Moreover it is interesting to note that a conditional model results in substantially reduced alphas. This suggests that part of what the Betting-Against-Beta portfolio offers comes from time-varying exposure to the 4-factors. A decomposition reveals that time-varying exposure to the market and value factors accounts for an economically significant portion of the Betting-Against-Beta premium.

Sefton et al (2011) use LN short-window regressions and the 3-factor model to investigate whether dynamic exposure explains the Betting-Against-Beta effect in the 500 European stocks with the highest market capitalization over the period 1994-2011. In this chapter I go further and use the 4-factor model on all NYSE/AMEX/NASDAQ stocks in the CRSP database, spanning 1926-2012. Furthermore, I explain the long-short portfolio alpha using a theoretical decomposition, and investigate the effect in subsamples 1926-1990 and 1990-2012. I identify that for the full CRSP cross-section, the effect disappears in the 1990-2012 subsample, coincident with a rise in leverage in the fund management industry.

II. Data

Market data was taken from CRSP, and covers all NYSE, AMEX and NASDAQ firms listed at any time between July 1926 till December 2012. In the case where any market data are missing, I substitute the previous available observation where appropriate. Daily returns are then calculated for each stock, using all available data on dividends, stock splits and delisting returns. In the event that a stock is delisted and no delisting return is available on CRSP then a return of $-30\%$ is used, following Shumway (1997). The risk-free rate, market, SMB, HML and WML factor return time series were taken from Kenneth French’s online data library.

Due to the breadth of the cross-section the sample includes a subset of stocks that are illiquid and have price observations that do not vary at the daily frequency. Such non-synchronous trading engenders the risk of understating betas, as in Epps (1979). However we need daily frequency observations in order to achieve convergent beta estimates that still remain pertinent to the risk of the stock at portfolio formation time. One way to correct for this is by excluding stocks with a price refresh rate of less than 75%. In fact for robustness three versions of
the study are run. V1 only considers stocks for inclusion in the sorted portfolios if they have 75% or more non-zero daily return observations over the six months leading up to rebalancing time. V2 and V3 only consider stocks if their market capitalization is above the 49th or 81st cross-sectional percentiles respectively, at rebalancing time. These thresholds were chosen because in 2012 the 49th and 81st percentiles of market capitalization ranked NYSE, AMEX and NASDAQ stocks in CRSP had market capitalizations of approximately $250M and $2bn, which correspond to the thresholds at which stocks might be classed as small-cap and mid-cap. Using the percentile as a filtering criterion rather than a nominal dollar value is robust to inflation over the 1926-2012 sample period.

III. Method

Every month I estimate the CAPM beta for each stock, using the prior 6 months of daily return observations. I then rank the stocks cross-sectionally according to their CAPM beta, and choose one portfolio consisting of the upper tercile of stocks (denoted “H”) and one portfolio consisting of the lower tercile of stocks (denoted “L”). Ideally quintiles would have been used, but during the early years of the sample period there were too few eligible firms to achieve sufficiently diversified quintile portfolios. Figure 26 shows that the number of eligible firms rises steadily over the years for V1, but nonetheless on the first portfolio formation date, 16th December 1926, there are only 365 firms that meet the criteria for inclusion in the study. The tercile sorts result in two different portfolios: H and L for the high beta and low beta stocks respectively. The equal-weighted and value-weighted daily returns on these portfolios are calculated for the ensuing month, before the monthly rebalancing occurs again. This is repeated every month from December 1926 till December 2012. Finally the returns on the H portfolios are subtracted from the returns on the L portfolios, to obtain two L-H return time series, one for the value-weighted case and one for the equal-weighted case.

FP suggests we should expect these L-H portfolios to exhibit a premium on the standard factor models in unconditional regressions. Here I intend to investigate whether this premium still exists in a conditional setting. Referring to Jagannathan and Wang (1996), if the CAPM holds conditionally but one estimates the coefficients using an unconditional OLS regression,
an extra term arises which biases the alpha estimate. This term is due to the covariance between the beta exposure and the market risk premium. To recap, here is the conditional CAPM:

\[ E[R_{i,t} | I_{t-1}] = \alpha_{t-1}^0 + MKT_{t-1} \beta_{MKT,i} \]

\( \alpha_{t-1}^0 \) is the conditional expected return on a "zero-beta" portfolio, \( MKT_{t-1} \) is the conditional market risk premium and \( \beta_{MKT,i} \) is the conditional beta of asset \( i \). Taking unconditional expectations of both sides:

\[ E[R_{i,t}] = \alpha^0 + MKT \bar{\beta}_{MKT,i} + cov(MKT_{t-1}, \beta_{MKT,i,t-1}) \]  

(3)

Where \( \alpha^0 = E[\alpha_{t-1}^0] \), \( MKT = E[MKT_{t-1}] \) and \( \bar{\beta}_{MKT,i} = E[\beta_{MKT,i,t-1}] \).

It is the third term in equation [3] that potentially causes bias. It arises due to the covariance between conditional risk premia and the portfolio’s exposure to the conditional risk premia, a phenomenon referred to in this chapter as "dynamic exposure". In Appendix 3A I derive an analogous result for the 4-factor model. One of the aims of this chapter is to establish whether the unconditional premium of the L-H beta-sorted portfolio is due to this bias. It is possible that the L-H portfolio offers no premium on conditional factor models. Given how central the linear relationship between beta and expected return is to financial theory, this study aims to determine whether the unconditional alpha achieved by the L-H beta-sorted portfolio on the 4-factor model remains significant in a conditional setting.

Cochrane (2005, page 145) summarizes the difficulty in reaching such an objective: "Models such as the CAPM imply a conditional linear factor model with respect to investors’ information sets. The best we can hope to do is test implications conditioned on variables that we observe. Thus, a conditional factor model is not testable!" The issue is severe. Yet LN’s short window regressions, using daily data and no conditioning variables, give direct estimates of conditional alphas and betas provided we are willing to believe that betas are relatively stable within the estimation window. Ang and Chen (2007, pages 10-12) derive the inconsistency bias and limiting distribution distortion for the unconditional OLS alphas and betas that result
if the true process generating stock returns follows a conditional CAPM. They specifically critique the short window regression method, as the bias and distortion terms they derive do not dissipate with higher sampling frequency, and moreover to correct for the bias and distortion would require parameters that can only be estimated using the true conditional beta time series. Since the conditional betas continue to vary within the short windows, the short window regression method does not capture the true conditional beta time series and it is not possible to make the correction. In response to this, LN (2006, Appendix B) use simulations to formally explore how high-frequency changes in beta affect their short-window regressions. They assume that both conditional betas and the conditional market risk premium follow AR(1) processes, that shocks to the conditional risk premium are allowed to covary negatively with market returns, and that shocks to conditional betas are allowed to covary negatively with stock returns. For simulations covering a wide range of empirically plausible parameter values, the high-frequency changes in beta do not impact substantially on conditional alpha estimation. In fact the short window regression technique works almost perfectly as long as market returns are not too highly correlated with shocks to the conditional risk premium. LN (2006)’s short window regression method is therefore favoured as a tractable route to estimating conditional factor coefficients.

I proceed by regressing the L-H portfolio time series on Carhart (1997)’s 4-factor model using unconditional OLS applied to the full-sample of data, from 1926-2012:

\[
R_t^L - R_t^H = \alpha + \beta_{MKT} R_{MKT} + \beta_{SMB} R_{SMB} + \beta_{HML} R_{HML} + \beta_{WML} R_{WML} + \epsilon_t \tag{4}
\]

Following this I apply LN’s short window regression method to estimate conditional 4-factor model coefficients. For each six-month window a standard OLS regression is run, resulting in a time series of conditional alphas for the L-H portfolio \(\{\alpha_t^c\}_{t=1}^T\). As well as running these regressions for the full-sample of data spanning 1926-2012, I additionally examine subsamples spanning 1926-1990 and 1990-2012. The choice of 1990 as the partition between subsamples is motivated by the growth in hedge fund AUM that occurred 1985-1994. Chany et al (2005) observe that, “the explosive growth in the hedge fund sector over the past several years has generated a rich literature both in academic and among practitioners.” Fung and Hsieh (1999,
Table 1) documents the "rapid growth in assets managed by an increasing number of hedge funds and CTA funds in the 1990s." Table 16 shows TASS data on hedge fund AUM taken from Fung and Hsieh (1999). AUM is expressed in absolute $ terms, as well as being expressed as a proportion of US GDP and as a proportion of the combined US$ GDP of the G7 countries. According to all three metrics hedge fund AUM grew considerably over the period 1985 to 1990, approximately by a factor of ten, and substantial growth in AUM continued throughout the 1990s. The exact choice of 1990 as the partition for the subsamples does remain somewhat arbitrary, but Tables 22 and 23 in Appendix 3B demonstrates that the main results of the study are robust if 1985 or 1994 are used instead.

Table 16: Number and Assets of Hedge Funds.
Hedge fund numbers and AUM are taken from the TASS database. GDP figures are taken from the World Bank. The aggregated G7 GDP figures were calculated by taking GDP observations for the US, UK, Canada, Japan, Germany, France and Italy, converting into US$ using contemporaneous exchange rates, and then summing.

<table>
<thead>
<tr>
<th>Year</th>
<th># Hedge Funds</th>
<th>AUM (US$ bn)</th>
<th>AUM / US GDP</th>
<th>AUM / G7 GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>37</td>
<td>0.4</td>
<td>0.009%</td>
<td>0.005%</td>
</tr>
<tr>
<td>1990</td>
<td>231</td>
<td>6.5</td>
<td>0.109%</td>
<td>0.044%</td>
</tr>
<tr>
<td>1991</td>
<td>310</td>
<td>10.1</td>
<td>0.164%</td>
<td>0.065%</td>
</tr>
<tr>
<td>1992</td>
<td>442</td>
<td>17.9</td>
<td>0.274%</td>
<td>0.107%</td>
</tr>
<tr>
<td>1993</td>
<td>644</td>
<td>35.8</td>
<td>0.520%</td>
<td>0.208%</td>
</tr>
<tr>
<td>1994</td>
<td>856</td>
<td>41.3</td>
<td>0.565%</td>
<td>0.225%</td>
</tr>
<tr>
<td>1995</td>
<td>1027</td>
<td>50.4</td>
<td>0.658%</td>
<td>0.252%</td>
</tr>
<tr>
<td>1996</td>
<td>1076</td>
<td>59.4</td>
<td>0.733%</td>
<td>0.298%</td>
</tr>
<tr>
<td>1997</td>
<td>987</td>
<td>64.6</td>
<td>0.750%</td>
<td>0.327%</td>
</tr>
</tbody>
</table>

Table 17 presents the regression results. Alphas are shown as annualized percentages. For each L-H portfolio I take the arithmetic mean of the time series of conditional 4-factor alphas, and calculate the percentage reduction this offers on the unconditional 4-factor alpha. T-stats for the alphas are shown in parentheses. Standard tests strongly reject null hypotheses of
homoskedasticity and zero autocorrelation in the unconditional OLS regression residuals, and therefore Newey-West HAC standard errors were used to calculate those $t$-stats. Lag length for the corrections was chosen optimally using the data, following Andrews and Monahan (1992). In all cases, the Newey-West standard errors result in smaller $t$-stats. For the conditional alphas $t$-stats are estimated using the sample variability of the conditional alphas as in LN (2006, Table 3). Note that because these standard errors are calculated using the sample variability of the conditional alphas, they should be robust to heteroskedasticity $^{13}$ and to autocorrelation $^{14}$.

IV. Results

For the full-sample the conditional alphas are lower than the unconditional alphas in an economically significant way, but the conditional and unconditional alphas remain statistically significant. However the subsamples appear to reveal a structural break in the data. While the L-H beta-sorted portfolio retains its statistically significant premium in the 1926-1990 subsample, this premium disappears in the 1990-2012 subsample. Furthermore for the later subsample, using a conditional 4-factor model as opposed to an unconditional model results in alphas that are up to 1004% smaller. For this subsample, dynamic exposure appears to be a substantial part of the L-H portfolio’s premium.

To test the hypothesis that there is a structural break in the vicinity of the year 1990 a Chow test is performed. The resulting $p$-values are given in Table 18. In every case the null hypothesis of parameter stability is unambiguously rejected, and we may conclude that a structural break occurs.

The results suggest that dynamic exposure to established sources of risk are a substantial part of the low risk anomaly in the full-sample and 1990-2012 subsample. The following

13Which does not affect the standard error of a sample average.
14Which shouldn’t exist in conditional alphas if the conditional 4-factor model holds since each alpha has a conditional mean of zero.
Table 17: Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta

Each month NYSE, AMEX and NASDAQ stocks are ranked according to their CAPM betas, as calculated using trailing 6 month daily returns. The lower tercile is chosen as portfolio L and the upper tercile is chosen as portfolio H. Daily returns are calculated on both portfolios in-between the monthly rebalancing. The returns on H are subtracted from the returns on L to obtain the dependent variable for the regressions: L-H. The regression specification is given by equation [4]. $\alpha^c$ is the mean of the conditional alphas resulting from LN (2006)'s short window regressions using 6-month intervals, and $\alpha^u$ is the standard unconditional OLS alpha. T-stats are given below in parentheses. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
</tr>
<tr>
<td>V1</td>
<td>2.68</td>
<td>4.49</td>
<td>40.36</td>
<td>4.09</td>
<td>5.05</td>
<td>18.98</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(3.59)</td>
<td></td>
<td>(3.69)</td>
<td>(3.72)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.55</td>
<td>8.09</td>
<td>19.10</td>
<td>7.12</td>
<td>8.09</td>
<td>12.00</td>
</tr>
<tr>
<td></td>
<td>(6.42)</td>
<td>(6.82)</td>
<td></td>
<td>(6.49)</td>
<td>(6.19)</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td>2.96</td>
<td>3.63</td>
<td>18.49</td>
<td>4.77</td>
<td>4.22</td>
<td>-12.82</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(2.92)</td>
<td></td>
<td>(4.06)</td>
<td>(3.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.76</td>
<td>6.70</td>
<td>14.08</td>
<td>6.72</td>
<td>7.04</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td>(5.94)</td>
<td>(5.69)</td>
<td></td>
<td>(5.79)</td>
<td>(5.56)</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td>2.31</td>
<td>2.90</td>
<td>20.46</td>
<td>3.97</td>
<td>3.64</td>
<td>-9.14</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(2.36)</td>
<td></td>
<td>(3.5)</td>
<td>(2.95)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.89</td>
<td>4.63</td>
<td>16.02</td>
<td>5.38</td>
<td>5.37</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>(4.22)</td>
<td>(3.80)</td>
<td></td>
<td>(5.34)</td>
<td>(4.53)</td>
<td></td>
</tr>
</tbody>
</table>
Table 18: Chow Test for a Structural Break between the 1926-1990 and 1990-2012 Sub-samples

The sample was split into subsamples spanning 1926-1990 and 1990-2012. For the full-sample and each of the subsamples the returns on the L-H portfolio were regressed on the 4-factor model. The residual sum of squares from these regressions were then used to compute the F-statistic for a standard Chow test. These F-statistics and the associated p-values are given below. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>vw</td>
<td>F-stat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>970.72</td>
<td>1100.51</td>
<td>1259.90</td>
</tr>
<tr>
<td>p-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ew</td>
<td>F-stat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1559.27</td>
<td>2524.14</td>
<td>2264.74</td>
</tr>
<tr>
<td>p-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

decomposition reveals which of these risk factors are contributing to the anomaly:

\[
\alpha_i^u = \alpha(\beta_i - \beta_i^u) + \text{cov} \left[ \text{MKT}_{t-1} \beta_{i,t-1}^{MKT} \right] + \ldots
\]

\[
\ldots \text{cov} \left[ \text{SMB}_{t-1} \beta_{i,t-1}^{SMB} \right] + \text{cov} \left[ \text{HML}_{t-1} \beta_{i,t-1}^{HML} \right] + \text{cov} \left[ \text{WML}_{t-1} \beta_{i,t-1}^{WML} \right] \quad (5)
\]

This decomposition is the analogue of LN (2006, equation 1), or Jagannathan and Wang (1996, equation 4), but which has been generalized to the 4-factor case. The term on the left-hand side is the alpha that arises theoretically if unconditional OLS is used to estimate coefficients when the true process generating returns is the conditional 4-factor model. Equation [5] decomposes the theoretical unconditional alpha into the linear sum of eight terms, where each of the 4 factors corresponds to two of those terms. The task is then to empirically estimate the terms from the right-hand side, so as to reveal how much of the L-H portfolio’s unconditional alpha is attributable to each of the MKT, SMB, HML and WML factors respectively. Please see Appendix 3A for the derivation, and a brief discussion on how the terms on the right-hand side can be estimated.
Table 19: Decomposition of the Unconditional Alpha

For each portfolio the quantities from the right-hand side of equation [5] were estimated empirically. Please refer to Appendix 3A. The portion of the unconditional alpha due to dynamic exposure to MKT is then given by $\hat{\alpha}_{\text{MKT}} \times (\hat{\beta}_{\text{MKT}} - \beta_{\text{MKT}}^{\text{MKT}}) + \text{cov}[\text{MKT}_{t-1}, \hat{\beta}_{\text{MKT}}^{\text{MKT}}]$, and this is recorded under the column MKT. Similarly for the other factors. $T = \text{MKT} + \text{SMB} + \text{HML} + \text{WML}$, and therefore $T$ equals the total unconditional alpha that should arise theoretically if the true process generating returns is a conditional 4-factor model. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MKT</td>
<td>SMB</td>
<td>HML</td>
<td>WML</td>
<td>T</td>
</tr>
<tr>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>1.68</td>
<td>0.06</td>
<td>1.48</td>
<td>-1.15</td>
<td>2.07</td>
</tr>
<tr>
<td>ew</td>
<td>1.61</td>
<td>-0.45</td>
<td>0.70</td>
<td>-0.04</td>
<td>1.82</td>
</tr>
<tr>
<td>V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>1.33</td>
<td>-0.41</td>
<td>1.32</td>
<td>-1.68</td>
<td>0.55</td>
</tr>
<tr>
<td>ew</td>
<td>1.39</td>
<td>-0.70</td>
<td>0.52</td>
<td>-0.26</td>
<td>0.95</td>
</tr>
<tr>
<td>V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>1.11</td>
<td>-0.24</td>
<td>1.15</td>
<td>-1.48</td>
<td>0.53</td>
</tr>
<tr>
<td>ew</td>
<td>1.38</td>
<td>-0.26</td>
<td>0.43</td>
<td>-0.88</td>
<td>0.67</td>
</tr>
</tbody>
</table>
Table 19 indicates that dynamic exposure to MKT and HML accounts for an economically significant portion of the unconditional alpha on the L-H portfolios. Dynamic exposure to WML appears to work in the opposite direction and would lower the unconditional alpha of the L-H portfolios, however its effect is dominated by MKT and HML. The effect of SMB depends on the subsample. $T$ is the total unconditional alpha predicted by the decomposition. Observe how the differences between unconditional alphas and mean conditional alphas in Table 17 correspond very closely to the unconditional alphas predicted in Table 19. Of course, the decomposition assumes that the 4-factor model holds conditionally and so will not explain any unconditional alpha that arises due to anomalies with the conditional 4-factor model.

In unreported results I analyse the statistical significance of each term from equation [5]. Standard errors for the covariances were estimated using the same method as in LN (2006, Table 6). The covariance terms were found to be statistically insignificant except for SMB, but on the other hand the mean conditional betas, unconditional betas and factor premiums were universally found to be highly significant.

V. Robustness Checks

When estimating CAPM betas on individual stocks the aim is to measure the systematic risk of the stock at portfolio formation time. Hence I use short 6-month windows for the regressions in order to capture conditional betas that are as close as possible to the underlying time-varying conditional beta process. Given the brevity of the window it is natural to choose daily frequency return observations to achieve convergent estimators. Unfortunately this leads to a potential issue, as non-synchronous trading for some stocks may result in beta estimates that are biased down. For example Epps (1979) investigates how correlations between major stocks in the same industry diminish as the frequency of return observations increases. Please

\footnote{Referring to Appendix 3A we use the following estimator: $\text{cov}(\hat{\beta}_{i,t}^{MKT}, MKT_t) = \text{cov}(\hat{b}_{i,t}^{MKT}, R_{Mkt})$. To obtain standard errors we first calculate $\sigma^2 = \text{var}(R_{Mkt})$ which is hence treated as a constant. We then run the following regression: $\hat{b}_{i,t}^{MKT} = \gamma_0 + \gamma_1 (R_{Mkt}/\sigma^2) + \epsilon_t$. Since $\gamma_1 = \text{cov}(\hat{b}_{i,t}^{MKT}, R_{Mkt})$, OLS gives us standard errors for the covariance estimator which implicitly condition on the sample variance of the conditional betas.}
see Barndorff-Nielsen et al (2008) for an up-to-date explanation of the non-synchronous data issue. I already filter out stocks with low price refresh rates (V1), or market capitalizations below the 49th or 81st percentiles (V2 and V3 respectively). Nonetheless as an additional control for non-synchronous data I repeat the experiment but using non-overlapping 3-day returns to estimate the CAPM betas on individual stocks. This still gives 42 observations per 6-month window. So it offers a good compromise between achieving convergence for the beta estimators, capturing conditional betas that are close to the underlying time-varying conditional betas, and alleviating potential issues with non-synchronous trading. In contrast weekly returns would only give 26 observations per 6-month window, and would be a poor alternative. Frazzini and Pedersen (2013) do something similar when estimating betas for individual securities, except that they use overlapping 3-day returns.

This repeat of the experiment is exactly the same as before, except that 3-day returns are used when estimating the CAPM betas on individual stocks. To calculate 3-day excess returns on the market portfolio I use the same method as described in Appendix 3A for calculating realized 6-month factor returns. Once the CAPM estimations are complete and the L and H portfolios have been chosen, the non-synchronous data issue effectively disappears. Hence as before I use the daily returns on the portfolio constituents to calculate daily returns on L and H, and run all of the 4-factor regressions using daily frequency return observations. Please see Table 20 for the results, which confirm that the conclusions of the study are robust. Furthermore note that LN (2006, Table 3) show how the mean conditional CAPM alphas estimated for SMB, HML and WML portfolios are fairly robust regardless of whether they use quarterly windows with daily returns, semi-annual windows with daily returns, semi-annual windows with weekly returns or annual windows with monthly returns. As a final robustness test, the experiment was rerun using quintile portfolios instead of terciles. Table 21 shows that the results of the study are also robust in this case.

VI. Conclusions

The low risk anomaly documented by Frazzini and Pedersen (2013) and Baker, Bradley and Wurgler (2011) is a paradoxical result that runs counter to an axiom of financial theory. In this chapter I demonstrate that the anomaly disappears in the 1990-2012 subsample, coincident
Table 20: Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta
Non-Synchronous Data

Each month NYSE, AMEX and NASDAQ stocks are ranked according to their CAPM betas, as calculated using trailing 6 month 3-day returns. The lower tercile is chosen as portfolio L and the upper tercile is chosen as portfolio H. Daily returns are calculated on both portfolios in-between the monthly rebalancing. The returns on H are subtracted from the returns on L to obtain the dependent variable for the regressions: L-H. The regression specification is given by equation [4]. $\alpha^c$ is the mean of the conditional alphas resulting from LN (2006)’s short window regressions using 6-month intervals, and $\alpha^u$ is the standard unconditional OLS alpha. T-stats are given below in parentheses. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
</tr>
<tr>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.67</td>
<td>4.62</td>
<td>42.24</td>
<td>4.08</td>
<td>5.38</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>(2.44)</td>
<td>(3.74)</td>
<td></td>
<td>(3.77)</td>
<td>(4.12)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>5.33</td>
<td>7.07</td>
<td>24.60</td>
<td>5.61</td>
<td>7.18</td>
<td>21.86</td>
</tr>
<tr>
<td></td>
<td>(5.88)</td>
<td>(6.57)</td>
<td></td>
<td>(5.49)</td>
<td>(6.02)</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.75</td>
<td>3.69</td>
<td>25.31</td>
<td>4.39</td>
<td>4.48</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(2.60)</td>
<td>(3.09)</td>
<td></td>
<td>(4.05)</td>
<td>(3.64)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>4.76</td>
<td>5.78</td>
<td>17.68</td>
<td>5.62</td>
<td>6.29</td>
<td>10.67</td>
</tr>
<tr>
<td></td>
<td>(5.18)</td>
<td>(5.33)</td>
<td></td>
<td>(5.19)</td>
<td>(5.37)</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.20</td>
<td>2.71</td>
<td>18.85</td>
<td>3.85</td>
<td>3.46</td>
<td>-11.22</td>
</tr>
<tr>
<td></td>
<td>(2.01)</td>
<td>(2.30)</td>
<td></td>
<td>(3.64)</td>
<td>(2.92)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>3.17</td>
<td>3.86</td>
<td>17.91</td>
<td>4.51</td>
<td>4.55</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>(3.49)</td>
<td>(3.41)</td>
<td></td>
<td>(4.77)</td>
<td>(4.03)</td>
<td></td>
</tr>
</tbody>
</table>
Table 21: Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta Quintile Portfolios

Each month NYSE, AMEX and NASDAQ stocks are ranked according to their CAPM betas, as calculated using trailing 6 month daily returns. The lower quintile is chosen as portfolio L and the upper quintile is chosen as portfolio H. Daily returns are calculated on both portfolios in-between the monthly rebalancing. The returns on H are subtracted from the returns on L to obtain the dependent variable for the regressions: L-H. The regression specification is given by equation [4]. $\alpha^c$ is the mean of the conditional alphas resulting from LN (2006)'s short window regressions using 6-month intervals, and $\alpha^u$ is the standard unconditional OLS alpha. T-stats are given below in parentheses. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
</tr>
<tr>
<td>V1</td>
<td>vw</td>
<td></td>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.73</td>
<td>6.14</td>
<td>39.17</td>
<td>(2.53)</td>
<td>(3.74)</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td>vw</td>
<td></td>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.51</td>
<td>5.60</td>
<td>19.44</td>
<td>(2.96)</td>
<td>(3.58)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.31</td>
<td>8.55</td>
<td>14.52</td>
<td>(5.98)</td>
<td>(5.73)</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td>vw</td>
<td></td>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.47</td>
<td>5.53</td>
<td>19.22</td>
<td>(3.74)</td>
<td>(3.54)</td>
<td></td>
</tr>
</tbody>
</table>
with a rise in hedge fund AUM. Furthermore, a conditional 4-factor model is capable of explaining an economically significant portion of the returns on the risk-sorted portfolio in the full-sample and the 1990-2012 subsample. In particular, it is time-varying exposure to the MKT and HML factors that contributes to the premium. One hypothetical reason for this structural break is that for the 1990-2012 subsample a smaller proportion of investors were constrained in their leverage, causing the security market line to steepen as low risk assets experienced increased demand, and high risk assets were sold short.
Appendix 3A: Alpha Decomposition

Assume that Carhart (1997)'s 4-factor model holds conditionally:

\[ E[R_{i,t}|I_{t-1}] = \alpha_{t-1}^0 + MKT_{t-1}\beta_{i,t-1}^{MKT} + SMB_{t-1}\beta_{i,t-1}^{SMB} + HML_{t-1}\beta_{i,t-1}^{HML} + WML_{t-1}\beta_{i,t-1}^{WML} \]

This is similar to Jagannathan and Wang (1996, equation 2) except that here we are using the 4-factor case. \( \alpha_{t,t-1}^0 \) is the conditional expected return on a "zero-beta" portfolio. Taking unconditional expectations of both sides we derive an expression for \( E[R_{i,t}] \):

\[ E[R_{i,t}] = E[\alpha_{t-1}^0] + E[MKT_{t-1}\beta_{i,t-1}^{MKT}] + E[SMB_{t-1}\beta_{i,t-1}^{SMB}] + ... \]

\[ ... + E[HML_{t-1}\beta_{i,t-1}^{HML}] + E[WML_{t-1}\beta_{i,t-1}^{WML}] \]

The average conditional excess return on the zero-beta portfolio is zero, i.e. \( E[\alpha_{t-1}^0] = 0 \).

From the definition of covariance \( E[XY] = \text{cov}[XY] + E[X]E[Y] \). Therefore:

\[ E[R_{i,t}] = \text{cov}[MKT_{t-1}\beta_{i,t-1}^{MKT}] + \text{cov}[SMB_{t-1}\beta_{i,t-1}^{SMB}] + \text{cov}[HML_{t-1}\beta_{i,t-1}^{HML}] + ... \]

\[ ... + \text{cov}[WML_{t-1}\beta_{i,t-1}^{WML}] + MKT\beta_{i}^{MKT} + SMB\beta_{i}^{SMB} + HML\beta_{i}^{HML} + WML\beta_{i}^{WML} \]

Where e.g. \( MKT \) is the average conditional premium, and \( \beta_{i}^{MKT} \) is the average conditional beta. Again this is a generalized 4-factor analogue of Jagannathan and Wang (1996, equation 4). Let’s define the following vectors:

\[ \alpha = (MKT, SMB, HML, WML) \]
\[
\beta_i = (\bar{\beta}_i^{MKT}, \bar{\beta}_i^{SMB}, \bar{\beta}_i^{HML}, \bar{\beta}_i^{WML})^T
\]

\[
\beta_i^u = (\hat{\beta}_i^{MKT}, \hat{\beta}_i^{SMB}, \hat{\beta}_i^{HML}, \hat{\beta}_i^{WML})^T
\]

Now we can derive something analogous to LN (2006, equation 1). An asset’s unconditional alpha is defined as \(\alpha_i^u = E[R_{i,t}] - \alpha \beta_i^u\). Therefore substituting for \(E[R_{i,t}]\) yields:

\[
\alpha_i^u = \alpha_i (\bar{\beta}_i - \beta_i^u) + \text{cov}[\text{MKT}_{t-1}\beta_i^{MKT}] + \ldots
\]

\[
\ldots \text{cov}[\text{SMB}_{t-1}\beta_i^{SMB}] + \text{cov}[\text{HML}_{t-1}\beta_i^{HML}] + \text{cov}[\text{WML}_{t-1}\beta_i^{WML}]
\]

Note that this is the same as equation [5]. It assumes that the true process generating returns is the conditional 4-factor model, and decomposes the resulting unconditional OLS alpha into eight terms, two for each of the 4-factors: (i) the difference between mean conditional beta and unconditional beta, scaled by the mean risk premium for that factor, and (ii) the covariance between the conditional premium and the portfolio’s conditional beta on that factor. Both of these quantities capture how dynamic exposure to any given factor contributes to the unconditional OLS alpha. This is intuitively obvious for the covariance terms, but also note that LN (2006, equations 1 and 2) show that the difference between the mean conditional beta and unconditional beta is related to dynamic exposure. The covariance terms can be estimated using LN (2006, equation 10):

\[
\text{cov}(\beta_i^{MKT}, R_{Mt}) = \text{cov}(\beta_i^{MKT}, R_{Mt}) = \text{cov}(\beta_i^{MKT}, MKT_t)
\]

Conditional betas estimated using 6-month short window regressions, \(\beta_i^{MKT}\), are used as proxies for the true conditional betas, \(\beta_i^{MKT}\), and the realized excess returns on the market over the 6-month windows, \(R_{Mt}\), are used as proxies for the conditional market premiums, \(MKT_t\).
Similarly for the other factors. To calculate the 6-month realized returns on the factors I first consider the long and short sides of the portfolio separately, and compound their daily return observations to reach their respective 6-month returns. I then subtract the 6-month return on the short position from the 6-month return on the long position to result in an accurate value for the 6-month return on the factor. For example with the SMB factor I calculate a time series of 6-month returns on the big (B) portfolio, and subtract that from the time series of 6-month returns on the small (S) portfolio. The average conditional premia, e.g. \( \bar{MKT} \), are calculated as the arithmetic mean of the 6-month realized returns. The remaining quantities that are required for equation [5] are all innocuous from an estimation perspective.
Appendix 3B: Alternative Subsamples
Each month NYSE, AMEX and NASDAQ stocks are ranked according to their CAPM betas, as calculated using trailing 6 month daily returns. The lower tercile is chosen as portfolio L and the upper tercile is chosen as portfolio H. Daily returns are calculated on both portfolios in-between the monthly rebalancing. The returns on H are subtracted from the returns on L to obtain the dependent variable for the regressions: L-H. The regression specification is given by equation [4]. $\alpha_c$ is the mean of the conditional alphas resulting from LN (2006)’s short window regressions using 6-month intervals, and $\alpha_u$ is the standard unconditional OLS alpha. $T$-stats are given below in parentheses. The year 1985 is used to partition the subsamples. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_c$</td>
<td>$\alpha_u$</td>
<td>% reduction</td>
</tr>
<tr>
<td>V1</td>
<td>vw</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.68</td>
<td>4.49</td>
<td>40.36</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(3.59)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.55</td>
<td>8.09</td>
<td>19.10</td>
</tr>
<tr>
<td></td>
<td>(6.42)</td>
<td>(6.82)</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td>vw</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.96</td>
<td>3.63</td>
<td>18.49</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(2.92)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.76</td>
<td>6.70</td>
<td>14.08</td>
</tr>
<tr>
<td></td>
<td>(5.94)</td>
<td>(5.69)</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td>vw</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.31</td>
<td>2.90</td>
<td>20.46</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(2.36)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ew</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.89</td>
<td>4.63</td>
<td>16.02</td>
</tr>
<tr>
<td></td>
<td>(4.22)</td>
<td>(3.80)</td>
<td></td>
</tr>
</tbody>
</table>
Table 23: Unconditional and Conditional 4-Factor Regressions for Portfolios Sorted on CAPM Beta (1994)

Each month NYSE, AMEX and NASDAQ stocks are ranked according to their CAPM betas, as calculated using trailing 6 month daily returns. The lower tercile is chosen as portfolio L and the upper tercile is chosen as portfolio H. Daily returns are calculated on both portfolios in-between the monthly rebalancing. The returns on H are subtracted from the returns on L to obtain the dependent variable for the regressions: L-H. The regression specification is given by equation [4]. $\alpha^c$ is the mean of the conditional alphas resulting from LN (2006)’s short window regressions using 6-month intervals, and $\alpha^u$ is the standard unconditional OLS alpha. T-stats are given below in parentheses. The year 1994 is used to partition the subsamples. vw and ew refer to value-weighting and equal-weighting for the L and H portfolios respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha^c$</td>
<td>$\alpha^u$</td>
<td>% reduction</td>
</tr>
<tr>
<td>V1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.68</td>
<td>4.49</td>
<td>40.36</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(3.59)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>6.55</td>
<td>8.09</td>
<td>19.10</td>
</tr>
<tr>
<td></td>
<td>(6.42)</td>
<td>(6.82)</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.96</td>
<td>3.63</td>
<td>18.49</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(2.92)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>5.76</td>
<td>6.70</td>
<td>14.08</td>
</tr>
<tr>
<td></td>
<td>(5.94)</td>
<td>(5.69)</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vw</td>
<td>2.31</td>
<td>2.90</td>
<td>20.46</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(2.36)</td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>3.89</td>
<td>4.63</td>
<td>16.02</td>
</tr>
<tr>
<td></td>
<td>(4.22)</td>
<td>(3.80)</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5
Conclusions of Thesis

This thesis has explored the relationship between economic fundamentals and stock prices via three empirical studies. I have presented new evidence for limited arbitrage in global equity markets, which is of major consequence to the task of measuring how economic fundamentals affect security prices. I have demonstrated a link between loadings on the Fama-French factors and firm-level default, which implies a relationship between priced aggregate macroeconomic risks and the financial performance of individual firms. Finally, I examined how frictions in financial markets alter the predictions of the most famous model in finance, the CAPM.

Since arbitrage is limited the implication is that despite the presence of rational arbitrageurs, sustained deviations from fundamental values can occur in security markets. Even if arbitrageurs agree on which model rationally prices securities as a function of economic fundamentals, there is no guarantee that the model’s predictions will hold empirically. Moreover, this fact makes it more challenging to identify the model in the first place. This establishes a caveat for any rational expectations model, and underlines the importance of research featuring behavioural traders, such as DeLong et al (1990a). The importance of behavioural traders to equilibrium prices could be one reason for the success of relative asset pricing models compared to absolute pricing models. Limited arbitrage also has consequence for the debate on financial regulation. It becomes harder to advocate an entirely unregulated free market if we acknowledge that markets are not efficient.

Nonetheless the evidence that loadings on the SMB and HML factors are related to incidence of firm-level default is consistent with a market where securities are rationally priced according to macroeconomic risks. As suggested by Petkova (2006), Lettau and Ludvigson (2001), Vassalou (2003) and Liew and Vassalou (2000), the SMB and HML portfolios are sensitive to news about the future state of the economy and innovations to the investment opportunity set. Here I demonstrate that stock returns are sensitive to these aggregate macroeconomic risks when their fundamental financial performance is poor and they are at risk of
bankruptcy.

We also see how frictions can decisively alter the role of fundamentals in asset pricing. The Betting-Against-Beta anomaly is so surprising because it runs completely contrary to the prediction of the CAPM, which has a strong theoretical motivation relating expected returns to economic fundamentals. The leverage-based explanation for Betting-Against-Beta is corroborated by my findings, and this confirms that any simplifying assumptions used in asset pricing theory must be tested. For example there is already a stream of literature which uses short-sales restrictions to explain a variety of empirical phenomena, including Hong and Sraer (2012), Ofek and Richardson (2003) and Scheinkman and Xiong (2003).

I have explored the complexity of the relationship between economic fundamentals and stock prices. The picture is one of security markets largely driven by aggregate macroeconomic risks that are priced efficiently by rational arbitrageurs, but where sustained deviations occur due to limited arbitrage and market frictions. Further research could focus on relaxing more of the idealized assumptions behind asset pricing models, as well as further incorporating behavioural phenomena into theoretical models, while attempting to balance parsimony and accuracy. It would also be interesting to do more work relating the financial performance of firms directly to the aggregate macroeconomic risk factors behind the SMB and HML portfolios.
Bibliography


Figure 1: V1 OOS Mispricing Episodes for Chile.
Figure 2: V1 OOS Mispricing Episodes for Denmark.
Figure 3: V1 OOS Mispricing Episodes for Spain.
Figure 4: V1 OOS Mispricing Episodes for Finland.
Figure 5: V1 OOS Mispricing Episodes for Hong Kong.
Figure 6: V1 OOS Mispricing Episodes for Italy.
Figure 7: V1 OOS Mispricing Episodes for Philippines.
Figure 8: V1 OOS Mispricing Episodes for Sweden.
Figure 9: V1 OOS Mispricing Episodes for Thailand.
Figure 10: Histogram of Price Refresh Rates.
The refresh rate is the proportion of non-zero daily return observations in the 250-day window used to estimate OLS 4-factor betas for a given firm-month. This histogram uses the pooled refresh rates for all firm-months in the unfiltered full-sample (D4).
Figure 11: Histogram of HML Betas for D1.

These histograms compare the winsorized $\beta_{HML}$ for the default firm-months vs. the non-default firm-months, using a sample that excludes low price shares (D1).
Figure 12: Histogram of HML Betas for D2.
These histograms compare the winsorized $\beta_{HML}$ for the default firm-months vs. the non-default firm-months, using a sample that includes low price shares (D2).
Figure 13: Histogram of Market Betas for D1.
These histograms compare the winsorized $\beta_{Mkt}$ for the default firm-months vs. the non-default firm-months, using a sample that excludes low price shares (D1).
Figure 14: Histogram of Market Betas for D2.
These histograms compare the winsorized $\beta_{Mkt}$ for the default firm-months vs. the non-default firm-months, using a sample that includes low price shares (D2).
Figure 15: Histogram of SMB Betas for D1.
These histograms compare the winsorized $\beta_{SMB}$ for the default firm-months vs. the non-default firm-months, using a sample that excludes low price shares (D1).
Figure 16: Histogram of SMB Betas for D2.
These histograms compare the winsorized $\beta_{SMB}$ for the default firm-months vs. the non-default firm-months, using a sample that includes low price shares (D2).
Figure 17: SVM Probability of Default for E1D1 as a Function of HML Beta.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that excludes low price shares (D1). This graph shows the SVM predicted probability of default as a function of $\beta_{HML}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 18: SVM Probability of Default for E1D1 as a Function of MKT Beta.

An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that excludes low price shares (D1). This graph shows the SVM predicted probability of default as a function of $\beta_{MKT}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 19: SVM Probability of Default for E1D1 as a Function of SMB Beta.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that excludes low price shares (D1). This graph shows the SVM predicted probability of default as a function of $\beta_{SMB}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 20: SVM Probability of Default for E1D2 as a Function of HML Beta.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). This graph shows the SVM predicted probability of default as a function of $\beta_{HML}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 21: SVM Probability of Default for E1D2 as a Function of MKT Beta.

An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). This graph shows the SVM predicted probability of default as a function of $\beta_{MKT}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 22: SVM Probability of Default for E1D2 as a Function of SMB Beta.

An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). This graph shows the SVM predicted probability of default as a function of $\beta_{SMB}$, holding the other explanatory variables fixed at their winsorized medians. ROC denotes the Receiver Operating Characteristic.
Figure 23: SVM Probability of Default for E1D2 as a Function of Weekly HML Beta.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). In this version, the 4-factor betas and SIGMA(e) were estimated using weekly returns, to correct for non-synchronous data. This graph shows the SVM predicted probability of default as a function of $\beta_{HML}$, holding the other explanatory variables fixed at their winsorized medians. *ROC* denotes the Receiver Operating Characteristic.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). In this version, the 4-factor betas and SIGMA(e) were estimated using weekly returns, to correct for non-synchronous data. This graph shows the SVM predicted probability of default as a function of $\beta_{MKT}$, holding the other explanatory variables fixed at their winsorized medians. $ROC$ denotes the Receiver Operating Characteristic.
Figure 25: SVM Probability of Default for E1D2 as a Function of Weekly SMB Beta.
An SVM was fitted using a Gaussian kernel and the full specification of explanatory variables (E1) applied to a sample of firm-months that includes low price shares (D2). In this version, the 4-factor betas and SIGMA(e) were estimated using weekly returns, to correct for non-synchronous data. This graph shows the SVM predicted probability of default as a function of $\beta_{SMB}$, holding the other explanatory variables fixed at their winsorized medians. ROC denotes the Receiver Operating Characteristic.
Figure 26: Number of Firms in Sample for V1.
This graph shows the number of NYSE, AMEX and NASDAQ firms in CRSP that have at least six months of market price observations, and 75% or more non-zero daily return observations over the six months leading up to time $t$. These criteria were used to choose the sample for V1, as a way to control for non-synchronous data issues.