

Idiosyncratic Risk and the Cross Section of Stock Returns

A thesis submitted for the degree of Doctor of Philosophy

by

Stanislav Bozhkov

Brunel Business School

Brunel University

November 2017

Abstract

A key prediction of the Capital Asset Pricing Model (CAPM) is that idiosyncratic risk is not priced by investors because in the absence of frictions it can be fully diversified away. In the presence of constraints on diversification, refinements of the CAPM conclude that the part of idiosyncratic risk that is not diversified should be priced. Recent empirical studies yielded mixed evidence with some studies finding positive correlation between idiosyncratic risk and stock returns, while other studies reported none or even negative correlation. In this thesis we revisit the problem whether idiosyncratic risk is priced by the stock market and what the probable causes for the mixed evidence produced by other studies, using monthly data for the US market covering the period from 1980 until 2013.

We find that one-period volatility forecasts are not significantly correlated with stock returns. On the other hand, the mean-reverting unconditional volatility is a robust predictor of returns. Consistent with economic theory, the size of the premium depends on the degree of ‘knowledge’ of the security among market participants. In particular, the premium for Nasdaq-traded stocks is higher than that for NYSE and Amex stocks. We also find stronger correlation between idiosyncratic risk and returns during recessions, which may suggest interaction of risk premium with decreased risk tolerance or other investment considerations like flight to safety or liquidity requirements. The difference between the correlations between the idiosyncratic volatility estimators used by other studies and the true risk metric – the mean-reverting volatility – is the likely cause for the mixed evidence produced by other studies. Our results are robust with respect to liquidity, momentum, return reversals, unadjusted price, liquidity, credit quality, omitted factors, and hold at daily frequency.

Acknowledgement

This thesis would not have been possible without the support of many people, to whom I am greatly indebted.

I wish to thank my supervisor Prof Habin Lee for the endless patience and Prof Sumon Bhaumik for his encouragement to embark on this study and to see it through despite all the odds.

Many thanks are due to Ms Vasilka Stamatova for her encouragement, support, optimism, and for proof-reading of this thesis (all errors that remain are solely mine).

Apologies are due to my immediate family for the edgy moods that occasionally accompanied my studies.

I am also grateful to my workmates who covered for my absences, often at very short notice.

Table of Contents

Abstract	2
Acknowledgement	3
1. Introduction	9
1.1. <i>Background to the work</i>	9
1.2. <i>Research motivation, aims and objectives</i>	14
1.3. <i>Structure of the thesis</i>	17
2. Related studies.....	19
2.1. <i>Introduction</i>	19
2.2. <i>Underlying economic theories</i>	21
2.3. <i>Empirical findings concerning idiosyncratic volatility and the cross-section of stock returns</i>	42
2.4. <i>Conclusions</i>	57
3. Research Methodology and Data Sources.....	59
3.1. <i>Introduction</i>	59
3.2. <i>Methodological notes</i>	60
3.2.1. <i>Research philosophy: positivism and its limitations</i>	60
3.2.2. <i>Deductive and inductive research</i>	63
3.2.3. <i>Quantitative and qualitative research</i>	72
3.2.4. <i>Statistical methods</i>	75
3.2.5. <i>Regression models</i>	79
3.2.5.1. <i>Correlation and causation</i>	79
3.2.5.2. <i>Factor model estimation</i>	80
3.2.5.3. <i>Selection of volatility models</i>	84
3.2.5.4. <i>Comparison of volatility forecasts</i>	93
3.2.5.5. <i>Assessing the correlation between expected idiosyncratic volatility and returns</i>	98
3.2.5.6. <i>Securities as assets</i>	101
3.3. <i>Control variables</i>	102
3.3.1. <i>Splitting total return into systematic and idiosyncratic returns</i>	102
3.3.2. <i>Control variables in the cross-section regressions</i>	107
3.4. <i>Data sources, transformations, and summary statistics</i>	108
3.4.1. <i>Data sources</i>	108

3.4.2. Calculated covariates.....	111
3.5. <i>Classification of volatility regimes</i>	122
3.6. <i>Conclusions</i>	128
4. Idiosyncratic Risk and the Cross-Section of Stock Returns: Empirical Findings	130
4.1. <i>Introduction</i>	130
4.2. <i>Comparison of volatility forecasts</i>	131
4.3. <i>Idiosyncratic volatility and the cross-section of stock returns: results from Fama–Macbeth cross-sectional regressions</i>	144
4.4. <i>Mean-reverting level of volatility</i>	157
4.5. <i>Further tests of robustness</i>	181
4.5.1. Was there an omitted factor?	181
4.5.2. Evidence from daily data	185
4.5.3. Portfolios as assets	191
4.5.4. Interaction effects	198
4.6. <i>Summary</i>	216
5. Discussion	217
5.1. <i>Introduction</i>	217
5.2. <i>Forecasts quality and goodness of fit</i>	217
5.3. <i>Mean-reverting level</i>	220
5.4. <i>Comparison with other studies</i>	222
5.5. <i>Liquidity premium and other premia</i>	227
5.6. <i>Is it tradable?</i>	232
6. Conclusions and directions for future research	236
6.1. <i>Conclusions</i>	236
6.2. <i>Limitations and directions for future research</i>	238

List of tables

Table 1: Summary of selected theoretical models and their predicted correlation between idiosyncratic risk and stock returns	40
Table 2: Summary of selected empirical studies of the link between idiosyncratic volatility and stock returns	56
Table 3: Deductive research workflow	71
Table 4: Average sector weights, Pearson correlations of sectors with the market, and cross-sector correlations	110
Table 5: Descriptive statistics, 1/1980–3/2013	113
Table 7: Mean volatilities by portfolios sorted by beta and capitalisation, 1980/7–2013/3	120
Table 8: Parameter estimates for Hidden Markov Model with two states and Normal distribution of volatilities in each state	124
Table 9: Parameter estimates for Hidden Markov Model with three states and Normal distribution of volatilities in each state	125
Table 10: Periods with high market volatility	126
Table 11: Average cross-sectional Spearman rank correlations of idiosyncratic volatilities....	133
Table 12: Dickey–Fuller tests for <i>ex ante</i> and <i>ex post</i> volatility estimates	136
Table 13: Predictive accuracy from Mincer-Zarnowitz regressions	140
Table 15: Fama–Macbeth cross-sectional regressions with historical and ARMA forecasts ...	154
Table 16: Fama–Macbeth cross-sectional regressions with mean-reverting volatility	160
Table 17: Descriptive statistics, 1/1980 - 3/2013	163
Table 18: Fama-Macbeth cross-sectional regressions with mean-reverting volatility – return persistence	178
Table 19: Summary statistics of the first common factor of idiosyncratic returns and loadings on that factor, 07/1982-03/2013	183
Table 20: Fama–Macbeth cross-sectional regressions with loading on the principal factor affecting idiosyncratic returns, 07/1982 – 03/2013.....	184
Table 21: Summary statistics of expected volatilities from daily data	187
Table 24: Portfolio Alphas Relative to Fama–French–Carhart Model	196
Table 26: Fama–Macbeth regressions with mean-reverting level of volatility – interactions with Roll's bid-ask spread, 7/1980-03/2013	206
Table 27: Fama–Macbeth regressions with mean-reverting level of volatility – interaction with	

unadjusted prices, 07/1980-03/2013	209
Table 28: Fama–Macbeth regressions with mean-reverting level of volatilities – interaction with traded volume in the last 36 months, 07/1980-03/2013	212

List of figures

Figure 1: Structure of Chapter 2.....	19
Figure 2: Efficient frontier with 30 stocks as of December 31, 2013, and a subset of 7 stocks ..	32
Figure 3: Volatility of DataStream market return index (all sectors)	123
Figure 4: Moving average and Exponentially-weighted moving average as filters.....	224

1. Introduction

1.1. Background to the work

Modern portfolio theory was developed in the middle of 20th century and over time became one of the thriving fields of economic research. In the span of less than ten years between the works of Markowitz and Lintner, it offered a new formulation of the decision-making problem in portfolio optimisation, which served as a basis for many subsequent theoretical enquiries and practical applications.¹

Markowitz (1952) proposed that rational investors in stocks should prefer higher expected return and lower variability of returns. This proposition was formalised in the concept of the efficient frontier, which was defined as the set portfolios offering lowest variance of returns for a given level of expected return.

The next significant step to solving the portfolio optimisation problem was made by Tobin (1956). He assumed that there existed one risk-free asset, usually referred to as cash. He demonstrated that investors could improve their expected return if they invested in a mix between the risk-free asset and some diversified portfolio on the efficient frontier. The returns obtainable through such a mix between risky and risk-free assets formed a straight line in the expected return/portfolio variance space, passing through the risk-free rate point and the selected efficient portfolio, and the steeper the line, the better the risk-return trade-off. Hence, the optimal solution was choosing a portfolio in which the line connecting the portfolio and the risk-free rate was tangent to the efficient frontier. In that setting all investors invested in a fraction of that super-efficient portfolio, irrespective of their risk appetite. The composition of that super-efficient portfolio still had to be estimated through the full Markowitz optimisation.

Sharpe (1964) and Lintner (1965b) developed the mean-variance reasoning further by formulating the Capital Asset Pricing Model, or “CAPM”. The CAPM demonstrated that

¹ Harry Markowitz shared the 1990 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel with William Sharpe and Merton Miller. He was praised in the award press release for “developing a rigorously formulated, operational theory for portfolio selection under uncertainty - a theory which evolved into a foundation for further research in financial economics”. Nobelprize.org, “The Prize in Economics 1990 - Press Release”, *Nobel Media AB 2014, 1990*
<http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1990/press.html>
[accessed 19 July 2016].

asset risk could be decomposed into two components – systematic risk and idiosyncratic risk. In their approach systematic risk referred to changes in asset prices that were due to common, market-wide changes that affected all securities, albeit to different extent, while idiosyncratic risk was the risk of change of individual asset prices due to reasons specific for the asset and uncorrelated with the overall market movements. Investors could diversify away all idiosyncratic risks by holding the market portfolio, defined as the value-weighted index of all financial instruments. Therefore, the CAPM predicted that asset returns in excess of the risk-free rate should be proportionate to the market excess return, and the coefficient of proportionality (referred to as asset's beta) should be determined by the covariance of the asset's excess returns with the market excess returns. In that context the idiosyncratic risk was the residual change of the values of individual assets that was uncorrelated with changes of the excess returns on the market portfolio, and that was assumed to be driven by random events concerning the individual issuer that did not have a broad market impact. The interpretation of the core CAPM proposition was then straightforward – only systemic risk was priced, idiosyncratic risk was not.

The conclusions of Sharpe and Lintner were surprising, yet intuitive and deceptively testable. However, the predicted optimal investor behaviour – holding a diversified portfolio of assets – was in stark contrast with surveys of the composition of individual portfolios. For example, Blume and Friend (1975) found that contrary to the predictions of CAPM, investors held rather concentrated portfolios. They examined a sample of tax returns from the 1971 tax year, and found as many as 34.1% of the tax returns listed only one dividend-paying stock², and 50.9% listed up to two dividend-paying shares.

Such apparently suboptimal behaviour could be at least explained by the restrictive assumptions of the CAPM, and economists were quick to explore the implications of the relaxation of some of these assumptions on equilibrium prices and returns. In particular, the CAPM assumed that there were no market frictions like transaction costs and asset indivisibility. Levy (1978) and Merton (1987) developed theoretical extensions of the CAPM that assumed that market imperfections prevented investors from investing in the entire available investment universe, and concluded that optimal investors would be seeking reward for the undiversified idiosyncratic risk.

Empirical studies to date, however, produced no conclusive evidence on the pricing of

² For tax purposes households were required to disclose only the dividend-paying stockholdings; information on the ownership of non-dividend-paying shares was not available in that sample.

idiosyncratic risk. The first tests of the CAPM performed by Lintner (1965a) confirmed that beta was a significant predictor of average asset returns, and the slope was positive, consistent with the theory. However, the coefficient for idiosyncratic risk turned out to be positive and statistically significant, contrary to the prediction of CAPM. Miller and Scholes (1972) also confirmed the findings of Lintner over an extended period and noted that idiosyncratic variance alone had higher correlation with average returns compared to beta alone. They also pointed out that the specification of the cross-sectional regression of realised returns on betas required the use of the true betas but in fact used the estimated ones, introducing error-in-variable problem.

The first tests of CAPM that addressed that error-in-variable problem were performed by Black et al. (1972) and by Fama and MacBeth (1973). Black et al. (1972) pooled securities into portfolios of similar securities in order to overcome the problem of the correlations between the pricing errors of individual securities. Fama and MacBeth (1973) further refined the method for testing the CAPM, and found that the idiosyncratic volatility was not a significant predictor of the cross-section of returns, which was consistent with the predictions of the CAPM and suggested that the previous positive results were the likely result of limitations of the testing methodology.

Since 2000, interest in the empirical investigation of significance of idiosyncratic risk in explaining the cross-section of stock returns rebounded. Such an interest was spurred by a variety of motivations. For example, the significance of beta as the sole, or at least a significant factor in explaining the cross-section of returns, has declined and some studies found it insignificant, while other market factors³ or characteristics of the issuer⁴ were added to empirical asset pricing models. The growing interest in behavioural asset pricing also contributed to the increasing interest in re-examining the assumptions of classical finance theory. Finally, the surge of studies in a related field – the correlation between idiosyncratic risk and aggregate market returns – also spilled over to the renewed interest in its contribution to explaining the cross-section of returns. Whichever the cause, in the past fifteen years a number of interesting studies were published that reported contradictory results concerning the significance of idiosyncratic risk as a predictor of expected returns.⁵

Malkiel and Xu (2004) found that CAPM beta was an important factor in explaining

³ e.g. the small-minus-big factor and the high- minus-low factors in Fama and French (1993)

⁴ e.g. size, price/earning ratio, dividend yield, past returns – see Daniel and Titman (1997, 1998)

⁵ Ang et al. (2006); Bali and Cakici (2008); Fu (2009); Huang, Liu, Rhee and Zhang (2012)

cross-sectional differences of returns but that its effect declined over time, but also confirmed that idiosyncratic risk was also a significant factor – both statistically and economically. Spiegel and Wang (2005) noted that there was a significant negative correlation between liquidity and idiosyncratic volatility, and that expected stock returns were positively correlated to idiosyncratic risk and negatively correlated to liquidity. They found that idiosyncratic risk was consistently positively correlated to expected stock returns, but the impact of one standard deviation change in idiosyncratic risk was on average between 2.5 and 8 times stronger than the impact of a corresponding one standard deviation increase in liquidity.

The debates were stirred by the surprising results of Ang et al. (2006) who reported that portfolios ranked on idiosyncratic volatility exhibited a consistently negative correlation to expected returns after controlling for various factors, across sub-samples, and for various portfolio formation strategies. Their results were puzzling because they contradicted both the CAPM (idiosyncratic volatility should not be priced at all), and the Levy and Merton models (idiosyncratic risk should earn positive risk premium). Ang et al. (2009) reviewed evidence from the G7 countries and other developed markets and found that the spread between the first and fifth quintiles of portfolios sorted on idiosyncratic risk was again negative, standing at -1.31 per cent per month after controlling for world market, and size and value factors.

Bali and Cakici (2008) also examined how idiosyncratic risk was priced using NYSE/AMEX/NADAQ data over the period July 1963 – December 2004 and found that the results of Ang et al. (2006) were not robust with respect to weighting scheme, time frequency, portfolio formation, and screening for size, price, and liquidity.

In a significant contribution to the debate, Fu (2009) observed that the theoretically correct variable to explain expected returns was the expected idiosyncratic risk in the current period, rather than the actual idiosyncratic risk in the preceding period, which was used by the studies of Ang et al. (2006) and Ang et al. (2009). He proposed to use the Exponential Generalised Autoregressive Conditional Heteroscedasticity (EGARCH) model in order to forecast next-period expected idiosyncratic volatility while allowing an asymmetric response to shocks. He used Fama and MacBeth (1973) with individual securities as assets and found that cross-sectional returns were statistically and economically significant and positively correlated with idiosyncratic risk. He found that reversals of returns were a significant factor for the puzzling results obtained by Ang *et al.*, (2006). Brockman et al. (2009) reached similar conclusions using an international data set and employing Fu's methodology. However, Guo et al. (2014) who argued that the way it was implemented involved a look-ahead bias that was

aggravated by the particular type of GARCH model used in the estimation (EGARCH(p,q) with $p, q = 1 \dots 3$). When they controlled for that bias they found that idiosyncratic risk was no longer significantly correlated with returns. In a similar vein, Huang, Liu, Rhee and Zhang (2012) argued that the omission of last-period return resulted in omitted variable bias, which could impact the significance of the idiosyncratic risk proxies.

Thus, the studies in the last fifteen years produced mixed evidence on the correlation between idiosyncratic risk and market returns. That topic, however, is of great importance from both a practical and a theoretical perspective. Finding the root cause for the mixed evidence produced by existing studies would allow improved portfolio construction and would optimise the risk-return trade-off for both individual and professional investors. For example, if idiosyncratic risk is correlated with returns, a failure to recognise it as a characteristic in portfolio construction could result in apparent over-performance of portfolios loaded on securities with higher idiosyncratic risk over portfolios with lower idiosyncratic risk; thus, hedge portfolios loaded on idiosyncratic risk could mislead investors into believing superior forecasting performance of their asset managers (high alpha) when in fact they are bearing higher idiosyncratic risk. On the other hand, if idiosyncratic risk is not priced by the market, then investment funds marketing portfolios constructed to exploit exposure to idiosyncratic risk could be overcharging investors and reducing their risk-adjusted performance after trading costs. Therefore, irrespective of the direction of the correlation, understanding its sign, magnitude of underlying cause should benefit investment managers.

The issue is also important from a theoretical perspective. Academic research has identified a number of stock market anomalies, as well as factors that seem to explain stock returns, but which lack a solid theoretical understanding. Understanding of the true role of idiosyncratic risk could allow improved return attribution and ultimately – improved identification of the underlying economic factors. For example, smaller stocks are known to have higher volatility; if idiosyncratic risk is priced by the market, then some of the documented variation of the small-minus-big factor could be due to changes of idiosyncratic volatility of small companies due to aggregate economic shock. Therefore, resolving the idiosyncratic volatility puzzle should benefit both practitioners and theorists and would further the cause of wealth management.

1.2. Research motivation, aims and objectives

In our view there are two important and closely interrelated aspects of this debate that deserve careful attention, and which are in the centre of this study: firstly, an analysis of the differences and similarities of the employed measures of idiosyncratic risk; secondly, analysis of the correlation between idiosyncratic risk and the cross-section of returns.

The first aspect concerns the use of different definitions of idiosyncratic volatility across studies. Thus, Malkiel and Xu (2004) measure idiosyncratic volatility as the standard deviation of residuals from the one-factor CAPM model or the three-factor Fama–French model. Ang et al. (2006) and Ang et al. (2009) use a similar approach but instead of applying it on a rolling window of monthly data, they use daily returns from the previous month in order to split volatility into systematic and idiosyncratic components and to calculate the mean daily variance in the respective month. Bali and Cakici (2008) employ two measures of idiosyncratic risk: the one used by Ang et al. (2006), and another version calculated for monthly data over the preceding 24 to 60 months, as available. Cao (2010) and Cao and Xu (2010) measure idiosyncratic volatility in terms of the exponentially-weighted moving average of the residuals from OLS regression (24 to 60 months) with weights of 0.9^k . Fu (2009), emphasising the importance of using forward-looking models in order to obtain *ex ante* risk measures, employs EGARCH(p,q) to produce forecasts from monthly data; a similar approach is followed by Spiegel and Wang (2005). Finally, Huang, Liu, Rhee and Zhang (2012) use forecasts from an ARIMA model fitted on realised monthly volatilities.

There is limited information on how these measures compare with one another in terms of forecasting the unobservable true volatility. Some of the studies that address the issue employ loss functions to compare idiosyncratic volatility estimates. However, loss functions could be of limited use when one compares estimates based on different frequencies (monthly vs daily). Bali and Cakici (2008), on the other hand, use Mincer and Zarnowitz (1969) regressions to compare predictive performance of monthly-based vs daily-based forecasts. Spiegel and Wang (2005) also perform a comparison of prediction accuracy but use only monthly data and loss-function approach to the test. Nonetheless, the choice of measure of true volatility may pre-determine the ‘winner’ in those forecasts.

Therefore, the first gap that this study aims to address is to compare the types of idiosyncratic volatility proxies employed by previous studies and to analyse how conclusions depend on the quality of source data.

Addressing the first gap would allow us to examine empirically whether idiosyncratic variance explains the cross-section of returns. Indeed, one should expect that if idiosyncratic risk explains the cross-section of returns, then the better forecasts would result in more robust prediction of returns. Furthermore, we can then analyse the role of other characteristics, e.g. the mean-reverting level of idiosyncratic volatility, in explaining the cross-section of returns. Such an analysis is warranted by the finding of Fu (2009) that idiosyncratic volatility is stationary for 90 per cent of all securities.

Another gap for our empirical study is to address some of the methodological challenges to the inspiring contribution of Fu (2009) – the alleged presence of look-ahead bias and the omitted-variable bias. Indeed, Huang, Liu, Rhee and Zhang (2012) report that introducing the lagged return as a control variable, which is omitted in other studies, including that of Fu (2009), renders idiosyncratic volatility statistically insignificant. Another criticism came from Guo et al. (2014) who argue that there might be look-ahead bias inherent in the studies employing EGARCH models, which explains the predictive performance of those forecasts.

In this study we aim to address those research gaps by addressing two interrelated research questions: firstly, to compare alternative volatility forecasts in terms of their accuracy, in order to understand if such differences explain the mixed empirical evidence in existing literature; secondly, to re-examine the relevance of under-diversification for asset prices by testing whether idiosyncratic risk is priced by investors in stock markets.

Therefore, the objectives of this research are as follows:

1. To compare the predictive performance of the main classes of estimators of monthly idiosyncratic volatility using Mincer-Zarnowitz regressions;
2. To test statistically whether the different estimates of idiosyncratic volatility explain the cross-section of monthly stock returns using the Fama-Macbeth methodology;
3. To test statistically whether other characteristics of the idiosyncratic volatility process, especially the mean-reverting level of volatility, explain the cross-section of stock returns using the Fama-Macbeth methodology.

The statistical tests use public domain secondary data on securities traded on stock markets in the United States of America over the period from January 1972 until March 2013. The source data is from the Thomson Reuters Datastream service. Most existing studies employ the dataset of the Center for Research in Security Prices (CERP), and in that respect the validation of the results from previous studies, in particular that of Fu (2009), against the Thomson Reuters dataset can be also construed as a test of whether previous results are

exposed to the data snooping problem.

The scope of this study is limited only to US stocks listed at the largest American stock exchanges during the period of study: the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the NASDAQ Stock Market (NASDAQ). The choice of market (the US) is aimed to facilitate comparisons with other existing studies in the field, while international (non-US) evidence is deferred for further study.

The underlying economic theory – the CAPM and its extensions by Levy (1978) and Merton (1987) – is not bound to a specific time frequency, and the results in principle should hold at all frequencies, e.g. daily, weekly, monthly, quarterly, or annual. Existing evidence on the significance of idiosyncratic risk is somewhat stronger with respect to daily frequency. However, there might be market microstructure effects that affect those findings. This is especially true for smaller stocks that have a narrower investor base and should earn higher return in equilibrium, according to the underlying economic model. Therefore, in this study we focus on the monthly frequency and leave the further cross-frequency validation for further studies.

In this dissertation we contribute to the debate in three ways. Firstly, we analyse how the different forecasts of the idiosyncratic variance perform in predicting next-period variance. The evidence available in that respect is scarce, and either compares only estimates from monthly data with one another, or uses proxies of true idiosyncratic variance that might be suspected to pre-determine the result of the comparison. Secondly, we re-examine the existing evidence concerning the link between idiosyncratic variances and the cross-section of stock return and we address the recent methodological objections concerning studies documenting the existence of a positive correlation. In that way we are able to bridge the gap between the quality of forecasts and the resulting explanations of the cross-section. We also address the criticisms (omitted variable and look-ahead biases) of the (E)GARCH methodology employed in some of the studies that find positive correlation between idiosyncratic risk and the cross-section of returns. Thirdly, we identify the mean-reverting level of volatility as a key variable in explaining the cross-section of returns. The outcome of our tests puts into perspective and reconciles the existing evidence by identifying the components of idiosyncratic volatility that explain the cross-section of returns.

Our results shall be useful from both a theoretical and a practical perspective. From the theoretical perspective evidence on the significance of idiosyncratic volatility would provide support to the Merton model. Furthermore, it would shed light on how significant that premium is, and indirectly – on whether under-diversification is widespread enough to affect

observed prices significantly. From a practical perspective the results could be used to construct investor portfolios and to assess their performance after taking into account their exposure to idiosyncratic risk.

1.3. Structure of the thesis

In Chapter 2 we review the findings of related studies. These studies fall into three broad categories, and we dedicate a section to each of those. In Section 2.2 we review the milestones of the modern portfolio theory until the formulation of the Capital Asset Pricing Model (CAPM). The CAPM predicts that idiosyncratic risk can be diversified away by investors completely, and therefore it is not priced. In particular, we see that there are at least two lines of extension, from which possible deviations from the predicted CAPM equilibrium could emerge. Firstly, this is the axiomatic basis of the Expected Utility theory. Criticisms from that direction are the basis of the behavioural asset pricing theory; such behavioural explanations resulting in aversion to idiosyncratic risk could be due to behavioural biases or regulatory requirements concerning the disclosure of significant individual loss-making positions. However, in this study we focus on concerns about the limits on portfolio diversification due to transaction costs, asset imperfect divisibility, or other motives that result in investors holding undiversified portfolios. In those cases the CAPM extensions of Levy (1978), Merton (1987) and Malkiel and Xu (2004) predict that idiosyncratic risk could be priced in equilibrium, if under-diversification is pervasive. In Section 2.3 we review the principal empirical findings concerning the idiosyncratic risk and the cross-section of returns.

Chapter 3 explains the principal methodological tools employed in this study. The reviews are fairly brief because we employ existing methods, so the chapter is more concerned with explaining the use of those methods in this study. In particular, Section 3.2 motivates the positivist philosophy that underpins the testing methodology of this study. Section 3.3 then explains how we split the observed excess return into a systematic component and an unexpected idiosyncratic return. The excess return is the variable that we aim to explain. It comprises of two components – systematic and idiosyncratic returns, and the latter are used to estimate idiosyncratic volatilities. Section 3.4 then explains how the idiosyncratic returns can be used to estimate idiosyncratic volatilities. In particular, we employ four such alternative estimators, and we dedicate a subsection to each of those.

Section 3.5 then explains our approach to comparing the predictive performance of these alternative volatility estimators. This comparison allows us to check how well each of the four estimators predicts returns, and which should shed light on whether it is indeed idiosyncratic risk that explains returns, and not some other factor or characteristic which correlates with some of the measure. Thus, if the superior estimators of volatility are significant predictors of the cross-section, then we could conclude that the noise in the inferior estimators is the reason for the mixed empirical performance. As it happens, it turns out that the opposite is the case, which suggests that there is another characteristic, that is more important than next-period expected volatility.

Section 3.6 lays out the Fama–Macbeth methodology that is used throughout Section 4 to perform the empirical tests. Section 3.7 then explains what data sources were employed in the empirical analysis, what transformations of those data were implemented, as well as how we classify the volatility states of the market, which is used in some of the robustness tests in Chapter 4.

Chapter 4 presents our empirical findings concerning idiosyncratic risk and stock returns. The structure of the chapter mirrors much of the structure of Chapter 3. In particular, Section 4.2 compares the volatility forecasts using the methodology of Section 3.5. Section 4.3 presents the results of the empirical tests concerning the correlation between idiosyncratic risk and returns using the Fama–Macbeth methodology described in Section 3.6. Section 4.4 presents further tests using the mean-reverting level of volatility as an explanatory variable instead of next-period volatility. The section also provides a first set of confirmatory robustness checks. These checks were further extended in Section 4.5, using additional tests that were not described in Chapter 3; the relevant methods are explained in each of the respective sub-sections.

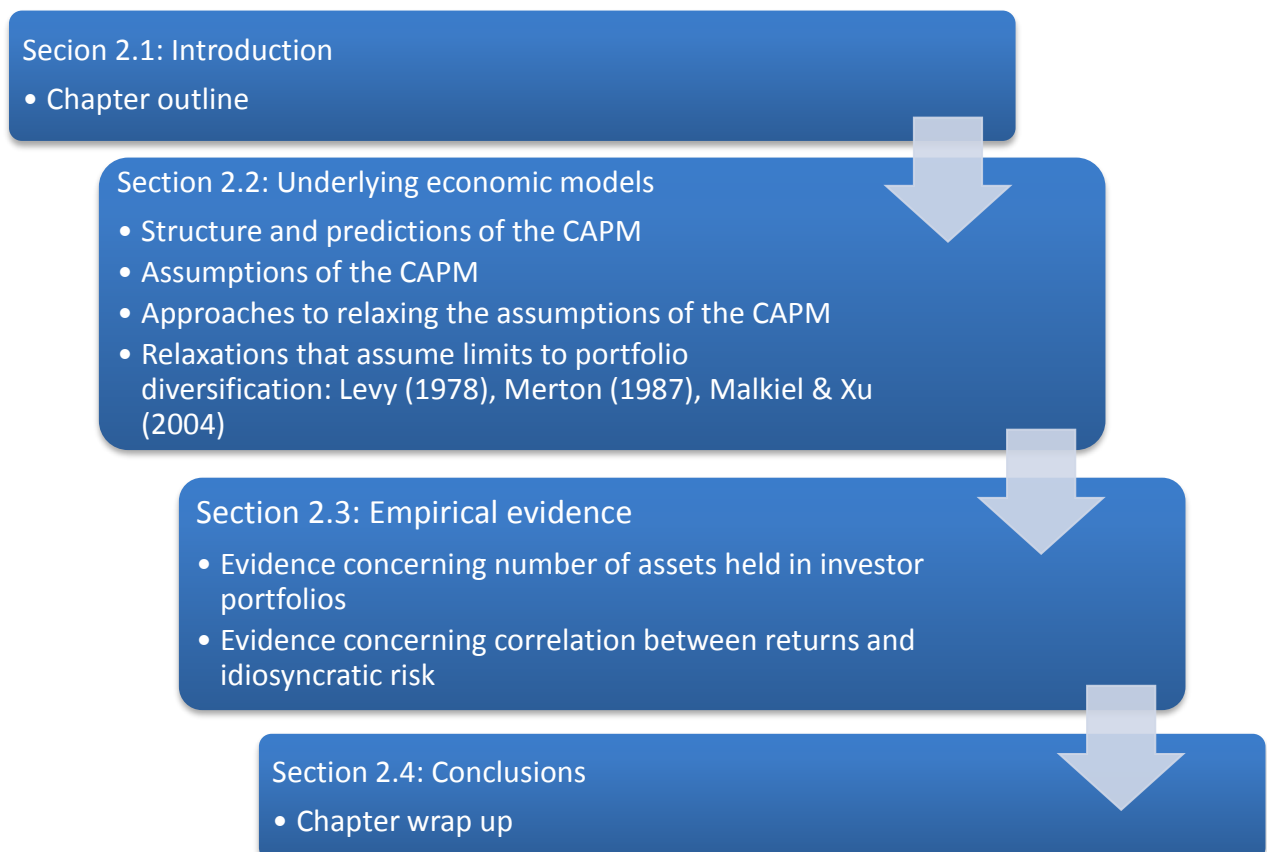
Chapter 5 discusses the results from our analysis, providing further motivations for the significance of the mean-reverting level of volatility, as well as the difficulties in interpreting the idiosyncratic volatility premium due to its inseparability from other characteristics, especially liquidity. The chapter also highlights the risk in practical implementation of trading strategies aiming to exploit the higher returns on high-volatility stocks. Chapter 6 concludes the thesis.

2. Related studies

2.1. Introduction

In this chapter we shall explore the theoretical foundations underpinning our study and the available empirical evidence. The body of related literature is quite vast, and this is related to the fact that the proposition that idiosyncratic risk is irrelevant for equilibrium prices is a core proposition of the classical finance theory. Therefore, in order to keep this chapter reasonably short and focused, we need to be very selective in our choice of related studies. Nevertheless, we also aimed to provide the readers with just enough context to enable them to relate our results not only with the results of classical finance theory, but also with the associated field of behavioural finance. Therefore, in Section 2.2 we shall outline the historical development of the Capital Asset Pricing Model and its refinements aimed to relax some of its assumptions. For convenience, we outline the structure of this chapter in Figure 1.

Figure 1: Structure of Chapter 2



Source: the author

The review of the CAPM serves two purposes: on the one hand, it outlines the key

model that predicts that idiosyncratic risk should not be priced by investors. On the other hand, the approach highlights the key assumptions underpinning that result. The key assumptions of the CAPM that are relevant for our topic are that investors' preferences can be represented by a quadratic utility function, and that there are no frictions. The simple statement of the first assumption partially conceals the true underlying assumptions, and we aim to highlight those by placing them in the appropriate historical context. In particular, we first need to assume that investor's preferences can be expressed in terms of a utility function, and then that function can be reasonably represented as a quadratic one. Furthermore, there should be no frictions like transaction costs or asset indivisibility.⁶ Relaxation of any of these assumptions could be relevant for our study. For example, relaxation of the expected utility theory axioms could give rise to a behavioural explanation of idiosyncratic risk.⁷ Absence of quadratic approximation to the utility function could imply significance of the higher moments of the utility function.⁸ Finally, in the presence of frictions that prevent full diversification, investors may require a risk premium for idiosyncratic risk.⁹ Therefore, we believe it is important to highlight how the predictions of the CAPM depend on those assumptions.

There are various relaxations that may give rise to deviations from the CAPM predictions. These can be broadly classified into behavioural extensions, which embed specific investor behaviours into the models in order to accomplish more realistic models of asset pricing. However, as we shall explain in somewhat greater detail in Chapter 3, these assumptions could be too *ad hoc*, which motivates our preference for the approach of classical finance to our problem. Therefore, we shall not explore behavioural asset pricing theories here; instead, in Section 2.2 we explore how the introduction of market frictions to the CAPM affects the predictions of that model. We review the models of Levy (1978), Merton (1987), and Malkiel and Xu (2004) in order to elucidate their assumptions and highlight their key predictions concerning the sign and the magnitude of such idiosyncratic risk premium.

Then, in Section 2.3 we review the available empirical evidence on the sign and significance of the link between idiosyncratic variance and stock returns. Such evidence broadly comprises two types of studies. Some studies observe investor behaviours (portfolio

⁶ See DeGenarro and Robotti (2007) for a discussion of financial market frictions.

⁷ For example, preference for own skewness, as proposed by Barberis and Huang (2008).

⁸ For example, Kraus et al. (1976) use a higher-order Taylor approximation of investor's utility function and conclude that there should be preference for co-skewness with the market.

⁹ Levy (1978); Merton (1987); Malkiel and Xu (2004)

compositions) directly, while most of the reviewed studies explore factors and characteristics that explain stock returns. Our review of those results has two goals: on the one hand, to understand the empirical support for the two competing theories (the CAPM and its extensions), and on the other hand to identify control variables that should be included in our study in order to avoid inasmuch as possible omitted variable bias. Finally, Section 2.4 sums up the arguments presented in this chapter.

2.2. Underlying economic theories

“Investing should be more like watching paint dry or watching grass grow. If you want excitement, take \$800 and go to Las Vegas.”

Paul Samuelson

“Wide diversification is only required when investors do not understand what they are doing.”

Warren Buffett

In 1713, Nicolas Bernoulli, brother of the famous mathematician Daniel Bernoulli, posed a problem in a letter to Pierre Raymond de Montmort. The problem later became known as the St Petersburg paradox and is stated as follows. Suppose that we are offered to bid in a lottery, the pay-off of which is determined by tossing a fair coin. The pay-off for the lottery starts at 1 ducat and is doubled every time heads appear. The first time tails appear, the lottery ends with the gambler winning the accumulated pay-off. So, if tails appear at the first toss, the gambler receives one ducat; if heads appear, the pay-off is doubled to two ducats. If at the second toss heads appear once more, the pay-off is doubled to four ducats and so forth. So, if k is the number of consecutive heads that appear, the pay-off of the bet equals 2^k ducats. The expected value of the bet would then be $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k 2^k = \sum_{k=1}^{\infty} 1 = \infty$. Hence the expected pay-off from playing this game is infinite. In contrast to the infinite expected pay-off, people are actually willing to pay a very modest entrance fee to bet on this game. This discrepancy between the finite (and in fact rather low) entrance fee people are willing to pay to play the game, and its infinite expected pay-off became known as the *St Petersburg paradox*.

Daniel Bernoulli published his solution to the puzzle in 1738 in the *Commentaries of the Imperial Academy of Science of Saint Petersburg*, where he argued that the price people are willing to pay to enter that lottery is determined by their evaluation of the likelihood of various outcomes and the utility those outcomes yield to them: “The determination of the value of an item must not be based on the price, but rather on the utility it yields. There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount” (quoted by Stewart, 2010: 26).

Bernoulli did not develop his thoughts on the *St Petersburg paradox* into a consistent theory. This was done by von Neumann and Morgenstern (1944) who developed the first coherent axiomatic theory of preferences under uncertainty. In their approach, economic agents chose between lotteries, where lottery u is defined as a list of possible states of the world, each occurring with a non-negative probability $p_k \geq 0$ and yielding a win of u_k at each state of the world; clearly $\sum_{k=1}^N p_k = 1, p_k \geq 0, \forall k$. They demonstrate that if there exists a preference relation that is complete, transitive, continuous, and independent, then there exists a utility function that represents these preferences. Thus, decision-making under uncertainty is reduced to the optimisation problem of maximising the expected utility subject to resource constraints. This result is known as the *Expected Utility Theorem* and bridges preference ordering to numerical utility values.

It is usually assumed that utility functions in money are continuous and monotonously increasing; the rationale is that more wealth is preferred to less wealth. The attitude towards risk is summarised by the sign of the second derivative of the utility function. Conventionally, a decision-maker is defined to be (strictly) risk-averse if receiving the expected value of a lottery is (strictly) preferred to partaking in the uncertain lottery. If he is indifferent between receiving the expected value of a lottery and the lottery itself, he is defined to be risk-neutral. It is known that an investor is strictly risk averse if and only if the utility function that represents their preferences is strictly concave and risk-neutral if the utility function is linear.¹⁰

There is an infinite number of utility functions that would satisfy the expected utility framework. Convenience and tractability, however, dictate the use of just a few utility functions that allow analytical tractability (ideally, closed-form solutions) of the implications of choosing a particular utility form. The three most commonly used utility functions are¹¹:

¹⁰ See proposition 6.C.1 in Mas-Colell, Whinston and Green, 1995: 187

¹¹ see Levy, 2012

1. Constant Absolute Risk Aversion (CARA) exponential utility function: $u(w) = -e^{-\alpha w}$;

2. Constant Relative Risk Aversion (CRRA) utility $u(w) = \frac{w^{1-\alpha}}{1-\alpha}$ for $\alpha \neq 1$, with limiting case the logarithmic utility $u(w) = \ln w$;

3. Quadratic utility function $u(w) = w - \frac{\alpha}{2}w^2$.

The CARA utility function owes its name to the fact that it is the only function for which the Arrow-Pratt risk aversion function is constant.¹² Normally one expects the absolute risk aversion $r_A(w)$ to be a decreasing function of wealth. This translates into decision-maker becoming more risk-tolerant, requiring smaller risk premium, with increase of wealth.

CRRA, on the other hand, represents the class of utility functions, for which relative risk aversion is constant.¹³

The third type of utility function listed above, the quadratic utility function, is straightforward to work with, and is the one that links the mean-variance optimisation framework with the expected utility theory. The downside of the quadratic utility is that it exhibits increasing absolute risk aversion, thus implying that decision makers should require a higher risk premium for identical lotteries with the increase of their wealth, which is a counter-intuitive prediction as wealthier investors should be more able to take risk and thus require a lower risk premium.¹⁴

¹² The Arrow-Pratt *absolute risk aversion* function (r_A) is given by $r_A = -\frac{u''}{u'} = -(\ln(u'))'$. The intuition behind that ratio is that it describes the curvature of the utility function around point w , while the metric remains invariant under linear transformations. Through Taylor expansion of the utility function it can be shown that the value of the risk premium that a utility maximiser would expect for taking part in a small lottery is given by $\pi \approx -\frac{1}{2}\sigma^2 \frac{u''(w)}{u'(w)}$ (Arrow, 1984: 151).

¹³ The relative risk aversion function (r_R) is defined as $r_R = -w \frac{u''}{u'}$. It measures the curvature of the utility function with respect to percentage changes of its argument, while the absolute risk aversion function quantifies that curvature with respect to absolute (dollar) changes of wealth. To see this consider the function $\tilde{u}(t) = u(tx)$ which links the utility of the initial wealth (at $t = 1$) to the utility of wealth increased by $(t - 1)\%$ to become tx . For fixed x , $\tilde{u}' = xu'(tx)$ and $\tilde{u}'' = x^2u''(tx)$, so that for $t = 1$ (the initial value of wealth) the absolute risk aversion becomes $\tilde{u}''(1)/\tilde{u}'(1) = xu''(x)/u'(x)$ (Arrow, 1984: 152). The requirement of decreasing absolute risk aversion of plausible utility functions is sometimes complemented by a requirement for non-increasing relative risk aversion.

¹⁴ Indeed, for the quadratic utility function, $r_A = -\frac{u''(w)}{u'(w)} = -\frac{\alpha}{1-\alpha w}$ and $r_A' = \frac{\alpha^2}{(1-\alpha w)^2} > 0$.

In an important contribution to the modern portfolio theory, Markowitz (1952) introduced the concept of mean-variance efficiency. He considers an investment universe of n assets. The returns on those assets are random variables; for convenience, returns here are defined as the ratio of the uncertain future price of the risky asset to the known (certain) current price of the asset. Let μ_i denote the expected return on asset i and let σ_{ij} denote the covariance between the returns on assets i and j . These portfolio returns can be summarised as a column vector $\mu = (\mu_1, \mu_2, \dots, \mu_n)'$, and a square covariance matrix $V = [\sigma_{ij}]$. It is also assumed that there are no redundant assets, i.e. assets the returns of which were a linear combination of the returns from other assets under any state of the world; i.e. the model excludes the possibility of rows (columns) in V that are linear combinations of other rows (columns). Also, the model assumes that there is no risk-free asset, i.e. there is no row (column) of zeros in V . By construction, V is square and symmetric, and the model assumes that it is also positive definite.¹⁵

Using the notation above, Markowitz defines variance-efficient portfolios as those having minimum variance of returns among those attaining certain minimum return, i.e.

$$\text{Min}_x \left\{ \frac{1}{2} Vx \mid \mu'x \geq \mu_p, 1'x = 1 \right\},$$

where x is the vector of portfolio weights (allocations) to each of the n available assets, and 1 is a vector of ones. Thus, an efficient portfolio allocation x minimises portfolio variance for some fixed expected return μ_p subject to investment of the entire wealth, i.e. $1'x = \sum_{i=1}^n x_i = 1$. A dual formulation is also possible, where investors chose weights x that maximise the expected return subject to the constraint that the variance of portfolio returns does not exceed the desired fixed threshold, i.e.

$$\max_x \left\{ x \mid \frac{1}{2} x'Vx \leq \sigma_p^2, 1'x = 1 \right\}.$$

Solving the variance-efficient portfolio allocation is thus reduced to solving a standard problem of quadratic programming. The Lagrangian for the problem is:¹⁶

$$\mathcal{L} = \frac{1}{2} x'Vx - \lambda_1(\mu'x - \mu_p) - \lambda_2(1'x - 1),$$

¹⁵ A matrix V is positive semi-definite if for every non-zero real column vector μ , $\mu'V\mu \geq 0$. If $\mu'V\mu > 0$, the matrix is positive definite. By construction, covariance matrices are positive semi-definite. Portfolio optimisation, however, requires the assumption of positive definiteness, which rules out perfectly correlated efficient portfolios. A positive-definite covariance matrix can be inverted, and that would allow unique determination of equilibrium prices.

¹⁶ The derivation outlined here follows the modern presentation of the mean-variance optimization problem rather than the original solution strategy of Markowitz.

and the corresponding first-order conditions (FOC) are

$$Vx = \lambda_1\mu + \lambda_2\mathbf{1},$$

together with the two original budget constraints. Hence the optimum weights are

$$x = \lambda_1V^{-1}\mu + \lambda_2V^{-1}\mathbf{1},$$

which could be substituted in in the two budget constraints, resulting in a system of two equations in two unknowns (the Lagrange multipliers); as long as V is positive definite, the system can be solved and so an optimal x exists. The system for the Lagrange multipliers has the form

$$\begin{aligned}\mu_p &= \lambda_1x'V^{-1}x + \lambda_2x'V^{-1}\mathbf{1}, \\ 1 &= \lambda_1\mathbf{1}'V^{-1}x + \lambda_2\mathbf{1}'V^{-1}\mathbf{1}.\end{aligned}$$

Since V is symmetric by construction, $x'V^{-1}\mathbf{1} = \mathbf{1}'V^{-1}x$ and the system can be written as

$$\begin{pmatrix} A & B \\ B' & C \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mu_p \\ 1 \end{pmatrix},$$

where $A = x'V^{-1}x$, $B = x'V^{-1}\mathbf{1}$, and $C = \mathbf{1}'V^{-1}\mathbf{1}$, so

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \frac{1}{AC-B^2} \begin{pmatrix} B & -C \\ -C & A \end{pmatrix} \begin{pmatrix} \mu_p \\ 1 \end{pmatrix}.$$

Substituting the solution for the Lagrange multipliers in the variance of the variance-efficient portfolio weights produces the formula for the efficient frontier:

$$\sigma_p^2 = \frac{A-2B\mu_p+C\mu_p^2}{AC-B^2},$$

which shows that the efficient portfolio is a parabola in the (μ_p, σ_p^2) space and hyperbola in the (μ_p, σ_p) (standard deviation) space.¹⁷

The mean-variance formulation of optimal investment could be interpreted as utility maximization with a quadratic utility function. That approach assumes that investors have quadratic utility function, i.e. $u(w) = w - \frac{\alpha}{2}w^2$. A rational investor seeks to maximise the expected utility, so

$$\begin{aligned}\mathbb{E}u(\mu) &= \mathbb{E}\left(\mu - \frac{\alpha}{2}\mu^2\right) \\ &= \left(\mathbb{E}\mu - \frac{\alpha}{2}\mathbb{E}(\mu^2)\right) \\ &= \left(\mathbb{E}\mu - \frac{\alpha}{2}(\mathbb{E}\mu)^2 - \frac{\alpha}{2}\text{var}(\mu)\right),\end{aligned}$$

¹⁷ Another useful result for variance-efficient portfolios is the *two-fund spanning* of the efficient frontier. It can be demonstrated that every variance-efficient portfolio can be constructed as a linear combination of two other efficient portfolios ('mutual funds') and thus the linear combination of any two efficient portfolios is also an efficient portfolio.

where $var(\mu)$ denotes the variance of expected returns. This shows that for an investor with quadratic preferences, for fixed variance of returns, the expected utility is maximised when the expected return is maximised.¹⁸ Alternatively, for a fixed level of expected return, the expected utility is maximised when the variance of the portfolio is minimised. Markowitz justifies the mean-variance optimisation framework as assuming a quadratic utility function (Markowitz, 1959, Chapter 13). He does not argue that investors' preferences are indeed quadratic but rather that the quadratic function could be a good local approximation to various plausible utility functions. He specifically considers logarithmic utility $\ln(1 + R)$, as well as quadratic approximation to $(1 + R)^{1/2}$ and $(1 + R)^{1/3}$.

Tobin (1956) and Tobin (1958) extend the portfolio model by adding a risk-free asset to the investment set, assuming that asset return distribution is a two-parameter distribution of the location-scale family. Under those assumptions Tobin shows that the problem of asset allocation is separate from the decision of risk tolerance. In particular, by definition the risk-free asset has no covariance with any of the risky assets, and since the covariance matrix for the risky assets is assumed to be positive definite, a risk-free asset could not be created through a combination (portfolio) of risky assets. Thus a portfolio comprising of the risk-free asset is variance-efficient among the portfolios with zero variance so it must belong to the efficient frontier. On the other hand, there must be another (risky) portfolio, also on the efficient frontier. Since all investors would be splitting their wealth between the risky asset and the risky fund, there would exist a risky portfolio that is efficient and all investors would be holding a share of it, hence by market clearance it must be simply the market portfolio comprising all risky assets in the economy. Just as in the case of the variance-efficient portfolio, the combination of the two efficient portfolios – the risk-free one and the market portfolio – spans the efficient frontier (two-fund spanning), and because the risk-free asset has no covariance with any of the risky assets, the efficient frontier transforms into a straight line through the market portfolio and the risk-free portfolio. In particular, the Sharpe ratio on any efficient portfolio c would equal the ratio for the market portfolio m , a result that became known as the *capital asset market line*, i.e. $(\mathbb{E}\mu_c - r_f)/\sigma_c = (\mathbb{E}\mu_p - r_f)/\sigma_p$, where r_f is the known (non-random) risk-free rate.

The work of Sharpe (1964) and the subsequent contribution by Lintner (1965b) formulate the Capital Asset Pricing Model (CAPM). Sharpe takes off from the observation of

¹⁸ Quadratic utility preferences make sense only in the increasing section of the function, hence higher $\mathbb{E}\mu$ guarantees higher $\mathbb{E}\mu - \frac{\alpha}{2}(\mathbb{E}\mu)^2$.

Markowitz (1959) that share prices tend to co-vary with the market, and demonstrated that what really matters for pricing is the systematic risk. The presentation of CAPM, as we know it presently, is formulated by Lintner (1965b), who shows that an investor's excess expected rate of return is related linearly to the risk of his total investment as measured by the standard deviation of his return.

An informal derivation of CAPM could follow the same variance-minimisation steps as above, but subject to a slightly modified budget constraint that adds a risk-free asset to the investment universe that yields a risk-free rate of return of r_f :

$$\min_x \left\{ \frac{1}{2} x' V x \mid (\mu - r_f \mathbf{1})' x \geq \mu_p - r_f \right\},$$

where $(\mu - r_f \mathbf{1})$ is the vector of excess returns. As before, one can form the Lagrangian and differentiate it with respect to portfolio weights x to obtain the first-order condition (FOC):

$$Vx = \lambda(\mu - r_f \mathbf{1}).$$

Let σ_{ie} be the covariance between some asset i (this could also be a portfolio of assets), and some efficient portfolio e , i.e. a portfolio that satisfies the FOC above. Then by definition $\sigma_{ie} = x_i' V x_e$, and upon substitution of the FOC, $\sigma_{ie} = \lambda x_i' (\mu_e - r_f \mathbf{1})$. This relation should hold for any efficient portfolio, and since the market portfolio m is an efficient one¹⁹, it follows that $\sigma_m^2 = \sigma_{mm} = \lambda(\mu_m - r_f)$. This gives the solution for the Lagrange multiplier λ and one can then eliminate it from the FOC formula above, giving the following link between the expected return of any asset i in relation to its risk as captured by σ_i and its covariance with the expected return on the market portfolio:

$$\mu_i - r_f = \beta_i (\mu_m - r_f), \quad (1)$$

where $\beta_i = \frac{\sigma_{im}}{\sigma_m^2}$. Therefore, in equilibrium the excess return on each asset would be

¹⁹ The market portfolio would be an efficient portfolio if the portfolios of the individual investors are on the efficient frontier. This follows because a combination of portfolios on the efficient frontier is also on the efficient frontier, and the market portfolio is the sum of individual investors' portfolios, which are assumed to be on the efficient frontier. In this informal outline we do not reproduce the proof of market clearance, but the issue of what happens when the portfolios of the individual investors are not on the efficient frontier is studied by Levy (1978), which we discuss in the following section. However, one should note that this general equilibrium solution is what sets out CAPM from all other models of asset returns: whereas other models simply presume the existence of some observed or unobserved factors driving returns and identify the arbitrage-free values of assets implied by exposures to those factors, the CAPM actually predicts what the sole factor of asset return is (the excess return on the market portfolio). Hence the special place of CAPM among the asset pricing models.

proportionate to the excess return of the market portfolio, and the coefficient of proportionality β_i equals the ratio of the covariance between the asset i and the market portfolio, divided by the variance of the market portfolio.

The predictions of CAPM are quite unequivocal in mathematical terms, yet the model's prediction might seem counter-intuitive. The CAPM denies the existence of abnormal returns. No matter how experienced the investment manager is, or what his cherished model for forecasting future returns is – technical, statistical, or a fundamental model – CAPM predicts that the expected return on his portfolio would depend solely on the betas of the assets in his portfolio and the share of investment in the risk-free asset.

Concerning the composition of portfolios of rational investors, CAPM makes two claims: firstly, all investors shall invest in identical portfolios of risky assets, with each portfolio having the same weights as the market portfolio. Differences in risk-tolerance of the different investors would be accommodated through changes in the share of investment in the risk-free asset. Risk-shy investors would invest heavily in risk-free assets, while investors seeking higher return would invest a larger share of their wealth in risky assets, and those seeking even higher returns could leverage their position by borrowing at the risk-free rate and investing in the market portfolio.

The important implication of CAPM that in equilibrium only the covariance of the stock with the market portfolio (systematic risk) is priced, while the idiosyncratic risk of the individual securities is not priced, is dependent on a number of explicit or implicit assumptions that are made in the derivation of the model: (i) investor preferences satisfy the axioms of von Neumann and Morgenstern and thus can be represented in terms of a utility function; (ii) investors have homogeneous beliefs; (iii) the utility function can be well approximated by a quadratic utility function; (iv) investors are price-takers; (v) the markets are frictionless and investor decisions are not affected by other concerns like taxes, transaction costs, asset indivisibility, asset liquidity; (vi) investors can borrow and lend in unlimited amounts at the market risk-free rate; (vii) all relevant information is available to all investors; (viii) markets are complete. This is indeed a long list of assumptions, and invited many subsequent studies. Some of these examine whether investor portfolios are sufficiently diversified, as the CAPM predicted that rational investors should seek a broad diversification of their asset holdings. Other studies seek to relax some of the assumptions and examine the impact of those relaxations on the market equilibrium. In the following paragraphs we briefly review the most relevant results for our study, while in a subsequent section we shall examine whether the CAPM performs well empirically despite the long list of restrictive assumptions.

The early presentations of the CAPM considered quadratic utility as a local approximation of some well-behaved utility function. That assumption is explored in more detail by Kraus et al. (1976), who employ a third-order approximation of the investor utility function, i.e.:

$$u(w) = u(\bar{w}) + \frac{u''(\bar{w})}{2!} \sigma(w) + \frac{u'''(\bar{w})}{3!} m_3(w) + O(w^4),$$

where $m_3(w)$ is the third moment of the random return process. Somewhat similar conclusions could be reached based on the prospect theory of decision-making under uncertainty, developed by Kahneman and Tversky (1979), which assigns an important role on the perceptions of potential gains and losses, rather than on expected return and standard deviation, and which could give rise to preference for lottery-like investments, characterised by higher skewness or kurtosis, and possibly – also higher idiosyncratic risk. Thus, Barberis and Huang (2008) build on the cumulative prospect theory and argue that the demand for lottery-like stocks (i.e. ones with significant positive skewness that have large growth potential) gives rise to the own skewness of the stocks being priced and stocks with high positive skew being overpriced.²⁰

The mean-variance framework underpinning the CAPM assumes that investors' preferences are fully captured in terms of their expected return and the variance of returns, so that the effects of higher-order terms of the Taylor expansion of the utility function, capturing skewness, kurtosis and higher moments could be ignored. This raises the problem in what cases could such an assumption be justified. This problem could be approached by asking what should be the joint distribution of stock returns that would allow investor preferences to be expressed only in terms of expected return and variance, irrespective of her utility function. The prime example of this situation is the case of normal returns because the Gaussian distribution is stable under addition²¹ and the distribution is entirely defined by its mean and standard deviation. In such a situation, the expected utility of an investor would be a function of the mean and standard deviation of the return distribution, irrespective of the functional

²⁰ Special care should be taken to distinguish the skewness preference in Barberis and Huang (2008) from that of Kraus et al. (1976), who predict preference for positive (co-)skewness based on Taylor approximation of the utility function of investors. The difference between the predictions of the two models is that Barberis and Huang (2008) predict significance of own skewness of the stocks, i.e. $\gamma(x) = \frac{\mathbb{E}(x - \mathbb{E}x)^3}{\sigma_x^3}$, whereas Kraus et al. (1976) predict significance of the co-skewness with the market, i.e. $\gamma(x, m) = \frac{\mathbb{E}(m - \mathbb{E}m)^2 \mathbb{E}(x - \mathbb{E}x)}{\sigma_m^3}$.

²¹ The sum of two normal variables is also normally distributed

form of the utility function. As long as the investor is risk-averse, for a fixed expected return she would choose the portfolio with the lowest variance; hence, if asset returns are jointly normally distributed, the optimal portfolio for the utility-maximising risk-averse investor would necessarily be a variance-efficient one.

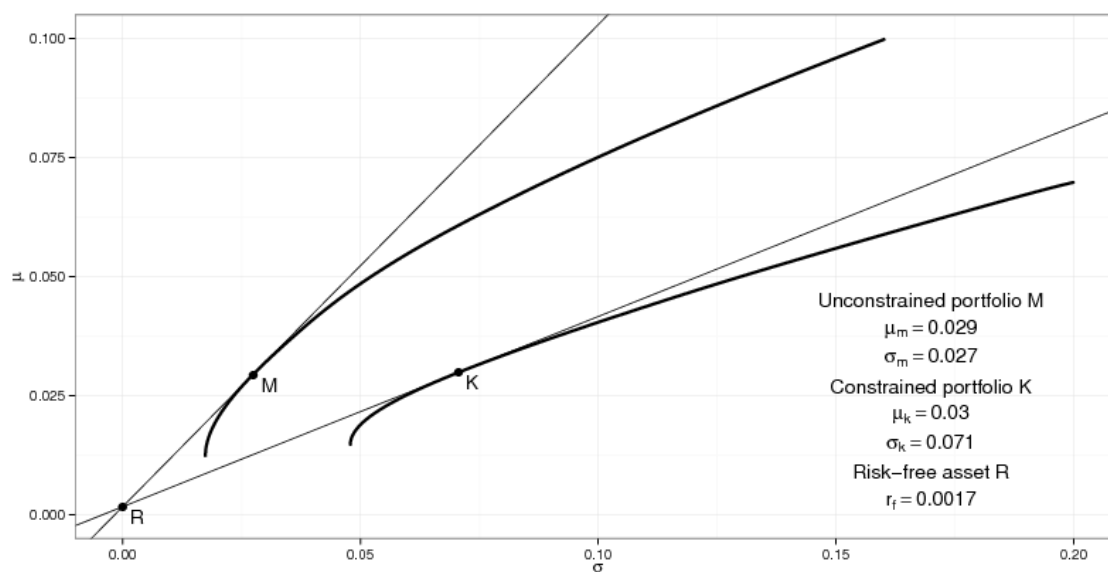
Chamberlain (1983) and Owen and Rabinovitch (1983) examine formally the circumstances in which investor's preferences could be formulated in terms of the mean and variance of portfolio returns alone. Chamberlain (1983) starts from the observation that if two portfolios yield the same expected utility under arbitrary utility function, then they should have the same distribution of returns and have coinciding mean and variance. Hence he establishes the conditions for the joint distribution of asset returns which would ensure that if the mean and variance of two portfolios coincide, then the portfolio return distributions are the same, i.e. the two portfolios would yield identical expected utility, irrespective of the utility function. Chamberlain demonstrates that an investor's utility function can be formulated in terms of mean and variance of returns if and only if returns are jointly elliptically distributed. Elliptic distributions are generalisation of the multivariate normal distribution. Formally they are defined in terms of the form of the distribution characteristic function. A multivariate distribution is said to be elliptical if its characteristic function is of the form $e^{it'\mu} \Psi(t'\Sigma t)$, where Σ is a positive semi-definite matrix, and μ is a vector of parameters. In the case of multivariate normal distribution, $\Psi(u) = e^{-u/2}$, and μ and Σ are the vector of means and the covariance matrix respectively. Linear combinations of elliptical distributions are elliptical too; the marginal distributions of a multivariate elliptical distribution are also elliptical. The class of elliptical distributions includes various symmetric distributions; in particular, it includes the normal distribution, Student's t, the logistic distribution, Laplace distribution, and symmetric stable distributions. While these distributions allow for a very broad set of return distributions that are characterised with varying tail fatness, all elliptical distributions have the limitation of being symmetric. The evidence concerning the joint distribution of returns is inconclusive. Levy (2012) surveys existing research and concludes that logistic distribution could not be rejected (as measured by the Kolmogorov-Smirnov test statistic) for daily, weekly and monthly returns for Dow Jones constituent stocks; other distributions that are reported as not rejected for certain stocks were beta, normal and log-normal distributions for monthly data, and Levy-stable distributions for the daily returns of some companies (Levy, 2012, Chapter 8). He also reports that at portfolio level and horizons up to nine months, logistic distribution cannot be rejected

at 10% confidence level, while on a 10- to 12-month horizon suitable distributions are beta, logistic, and Weibull. The evidence of Levy provides solid ground to consider the mean-variance framework as the baseline case of modern finance theory. It is clear that such evidence depends on the set of investigated distributions, the list of considered assets, as well as the employed goodness of fit statistic. However, there are other studies, e.g. Chicheportiche and Bouchaud (2012), that reject the joint elliptical distribution of returns.

The CAPM assumes a frictionless market where assets are perfectly divisible, there are no transaction costs for selling, buying or holding securities, and no liquidity differences. These are strong assumptions, and both researchers and practitioners alike have sought to relax those assumptions and to shed light on the structure of the resulting market equilibrium. Thus, Levy (1978) poses an identical variance-minimisation problem as that solved by Markowitz and CAPM. However, he assumes that for some reason each investor invests in some subset of the investment universe. The number of securities for the k -th investor was n_k , while the total number of shares in the market was n , $n_k \leq n$. Levy assumes that each investor can invest in some n_k pre-determined securities and the risk-free asset.²² This means that in mathematical terms, the optimisation is identical to CAPM, except that the investment decision for investor k involves not the market efficiency frontier (the efficiency frontier that was feasible when each of the all n assets were used in portfolio), but the efficiency frontier spanned by using the n_k assets, in which investor k could invest. Thus the efficiency frontier attainable by investor k is in the interior of the market efficiency frontier, and the capital allocation line for investor k is below the capital allocation line for the unconstrained investors (those who could invest in all assets).

²² Selecting an optimal portfolio subject to cardinality constraint, i.e. selecting n_k shares out of n available shares, is a hard problem of combinatorial optimisation.

Figure 2: Efficient frontier with 30 stocks as of December 31, 2013, and a subset of 7 stocks



Source: author's calculations

This point is illustrated on **Error! Reference source not found.** above. There we use 30 large stocks (approximately those of Dow Jones Industrial Average) to calculate the efficiency frontier as at the end of 2013. The risk-free portfolio almost coincides with the origin, as the risk-free rate of return (in that case: one-month constant maturity US treasury bills) is just 0.02% p.a. The higher curve on that plot shows the efficiency frontier that could be reached by investing in all of the 30 assets. For the purposes of the example, we consider that these 30 DJIA stocks are all the available securities in the market. Then the tangency portfolio is given by the point M . According to CAPM, each investor in this economy would hold some combination of the risk-free asset R and the market portfolio M . Thus, the optimal allocations for investors (the capital market line) would be the line RM . Should some investors be willing to earn higher return than M , they should leverage their position, i.e. borrow money at the risk-free rate r_f and invest in the market portfolio M . Thus, RM is the capital market line to the unconstrained investors.

Then we consider an investor who prefers another investment strategy: the ABC strategy. This strategy reasons as follows: holding all 30 stocks is impractical because of costs for maintaining and rebalancing the portfolio. Instead, that investor chooses to invest only in stocks with tickers starting with letters A, B, or C. This could be a practical strategy in its idiosyncratic way because one could find the price quotes in the newspaper easily – they

would be on the first page of market data. So in this particular case our investor is limited to investing only in the following stocks: AA (Alcoa Inc), AXP (American Express), BA (Boeing), BAC (Bank of America), CAT (Caterpillar), CSCO (Cisco Systems), and CVX (Chevron). This is not too bad a strategy as the stocks are from quite diverse sectors of the economy. Also, the expected return upon investment only in the ABC-stocks (portfolio K) even slightly outperforms portfolio M in terms of expected return. Unfortunately, the risk taken by the ABC investors is 2.6 times higher compared to the market risk as measured by the standard deviations of returns for portfolio M .

Solving the same problem as the CAPM in the previous section, the ABC investor would be in equilibrium if

$$\mu_i - r_f = (\mu_k - r_f) \frac{\sigma_{ik}}{\sigma_k^2}.$$

Levy points out that this equilibrium condition is very different from that implied by CAPM: the market price of risk ($= (\mu_k - r_f)/\sigma_k^2$, i.e. the slope of the capital market line) is measured relative to the co-variation of asset i with the constrained portfolio K instead with the unconstrained portfolio M . The covariance of asset i is also measured relative to portfolio K rather than the market portfolio M . Thus, investors in the same security i face different prices of risk and covariance with the constrained portfolio that depended on what other assets the investor has selected for his portfolio. For example, the securities of Alcoa are included in both portfolio K and portfolio M , yet for portfolio K the Sharpe ratio (the price of risk) was 0.4 compared to 1.0 for portfolio M . This plurality of individual variance-efficient frontiers obscures its implications for the equilibrium prices. However, Levy (1978) demonstrates that if investors hold very undiversified portfolios, then expected returns in equilibrium would depend on idiosyncratic risk. At the same time, betas with the whole market could be irrelevant because they would include many covariance terms with other securities, which investors actually may not hold in their portfolios. One way to look at this could be to note that the covariance of returns of a portfolio with n assets and returns μ_n would by definition equal the sum of n individual variance terms plus $(n^2 - n)$ covariances between individual assets. When n is large, the variance of the market portfolio would depend on the covariance terms and the impact of individual variances on portfolio variance would be smaller; with a smaller number of assets in the portfolio, the impact of individual variances on portfolio variance could not be neglected. Upon aggregation of the individual equilibrium conditions, Levy demonstrates that in his model the market price of risk from the CAPM

model $(\gamma)^{23}$ is not applicable. Instead, he finds that the market price in the constrained model (γ_1) equals $\gamma_1 = \gamma \frac{\sigma_m^2}{\sum_k t_k \sigma_k^2}$, where the sum in the denominator runs over the optimal portfolios for each investor k , and t_k is the share of the risky portfolio of investor k in the total market. This leads him to conclude that “the classic CAPM may be the approximate equilibrium model for stocks of firms which are held by many investors (for example, AT&T), but not for small firms whose stocks are held by a relatively small group of investors.” (Levy, 1978: 650).

In terms of its implications for empirical testing, the model developed by Levy (1978) highlights the difficulties involved in such tests. In particular, allocation decisions are based on betas with some unobservable portfolios that are different from the market portfolio. Therefore, beta with market portfolio needed not be a good measure of the risk faced by the under-diversified investors. In fact he shows that the betas for constrained investors would be higher than those in the unconstrained CAPM world, and hence the average beta across all investors would also be strictly higher than the CAPM beta. In the absence of a good proxy of true betas, Levy (1978) reasons that idiosyncratic variances could be expected to be superior predictors of the price behaviour than the traditional beta. (p. 654)

Merton (1987) develops a similar model but due to the somewhat different structure of the model, he is able to draw more specific conclusions about the size of the premium attributable to idiosyncratic risk. His model is again one of a two-period economy where some of the investors ‘know’²⁴ only a proper subset²⁵ of the universe of available securities. In the complete information case where all investors know all securities, the model reduces to the CAPM. In Merton’s model some investors may know all securities, but at least some of the investors are assumed to ‘know’ only some of the securities. The ‘knowledge’ of a security in that model is perfect – if an investor was ‘informed’ of a security, he or she knows the return generation process for that security and agrees with all other investors who know that security on the parameter values for that security (conditional homogeneous beliefs). The key behavioural assumption in the model is that investor would include a certain security k in his

²³ γ , also called Sharpe ratio, equals to excess return on the market portfolio divided by the standard deviation of the market portfolio, i.e. $\gamma = (\mu_m - r_f)/\sigma_m$.

²⁴ ‘Knowing’ here has essentially the same meaning as in the work of Levy (1978), i.e. if an investor ‘knows’ a security, it means that he has somehow pre-selected that security for his portfolio. The model makes no assumptions as to how investors came to know or chose to know some security.

²⁵ A subset is said to be proper if there are elements of the superset that are not in the subset.

portfolio optimisation only if the investor ‘knows’ that security. Merton (1987) motivates that assumption as follows (op. cit., p. 488):

“The prime motivation for this assumption is the plain fact that the portfolios held by actual investors (both individual and institutional) contain only a small fraction of the thousands of traded securities available. There are, of course, a number of other factors (e.g., market segmentation and institutional restrictions including limitations on short sales, taxes, transactions costs, liquidity, imperfect divisibility of securities) in addition to incomplete information that in varying degrees, could contribute to this observed behavior. Because this behavior can be derived from a variety of underlying structural assumptions, the formally-derived equilibrium-pricing results are the theoretical analog to reduced-form equations.”

The model assumes a two-period economy comprising n firms. The shares of those firms are traded at the beginning of the period. The cash flows generated from investing in those firms consist of three components – an expected return on the investment, an economy-wide shock, the exposure (loading) to which is company-specific, and idiosyncratic shock for each company. The economy-wide shock in this model specifies the correlation structure between the returns of the different companies.²⁶ If the starting investment for company k is I_k , then the end-of-period cash flows (\tilde{C}_k) are given by

$$\tilde{C}_k = I_k[\mu_k + a_k\tilde{Y} + s_k\tilde{\varepsilon}_k], \quad (2)$$

where μ_k , a_k and s_k are fixed company-specific parameters describing the production technology of company k . \tilde{Y} is assumed to be a random variable, describing the state of the world at the terminal period. It is a common factor reflecting the uncertainty in the model and is the common factor underlying the cash flows of all companies $k = 1 \dots n$. It is assumed that it has zero mean and unit variance, i.e. $\mathbb{E}(Y) = 0$ and $\mathbb{E}(Y^2) = 1$ (in order to avoid adding indices 0 and 1 to starting and end-of-period values we follow the notation of Merton (1987) and use tilde to mark realised, end-of-period values, and bar to mark the expected values). $\tilde{\varepsilon}_k$ represents the idiosyncratic risk in the model, i.e. shocks to the

²⁶ Let two random variables X and Y depend on a common random variable Z and independent shocks ε_x and ε_y , i.e. $X = \beta_x Z + \varepsilon_x$, and $Y = \beta_y Z + \varepsilon_y$. Then, $(X - \beta_x Z)$ and $(Y - \beta_y Z)$ are independent, i.e. $\mathbb{E}(X - \beta_x Z)(Y - \beta_y Z) = 0$, so that $cov(X, Y) = \beta_x \beta_y \sigma_Z^2$.

end-of-period cash flows that are specific to firm k and do not depend on either the idiosyncratic shocks to the rest of the firms, or on the state of the world (\tilde{Y}), i.e. $\mathbb{E}(\tilde{\varepsilon}_k | \tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_{k-1}, \tilde{\varepsilon}_{k+1}, \dots, \tilde{\varepsilon}_n, Y) = 0, \forall k$.

Let V_k denote the equilibrium value of company k at the initial point, and let \tilde{R}_k be the realised return per dollar from investing in company k . Since this is a two-period model, all of the cash flows of company k are distributed to shareholders at the end of the period. Hence, $\tilde{R}_k \equiv \tilde{C}_k/V_k = \bar{R}_k + b_k \tilde{Y} + \sigma_k \tilde{\varepsilon}_k$, where $\bar{R}_k = \mathbb{E}(\tilde{R}_k) = I_k \mu_k / V_k$, $b_k = a_k I_k / V_k$, and $\sigma_k = s_k I_k / V_k$. Besides the n shares, the economy also has a risk-free asset, the return on which is denoted by R (note that R denotes one plus the risk-free rate r_f , $R \equiv 1 + r$), and one other composite asset that combines the risk-free asset with a forward contract on the state of the world Y ,²⁷ the return on which is given by $\tilde{R}_{n+1} = \bar{R}_{n+1} + \tilde{Y}$ (the index $n + 1$ shall be used to denote that composite security). The risk-free and the composite securities are assumed to be interior to the efficient frontier so that investment in these securities is zero.

Investors are assumed to be risk-averse. Each investor j is endowed with an initial endowment W^j that he invests in a portfolio composed of some or all of the $(n + 2)$ securities traded in the market at the initial date. The return on that portfolio is \tilde{R}^j , so the wealth at the terminal period is $\tilde{W}^j = \tilde{R}^j W^j$. The investor preferences are assumed to be quadratic and specified as $U_j = \mathbb{E}(\tilde{W}^j) - \frac{\delta_j}{2W^j} \text{Var}(\tilde{W}^j)$. For the latter parts of the derivation it is also assumed that investors have identical risk preferences (i.e., $\delta_j = \delta, \forall j$) and initial endowments ($W^j = W, \forall j$).

In this setting Merton demonstrates that the resulting equilibrium value \tilde{V}_k of firm k equals the equilibrium value V_k^* when all investors are informed of security k , discounted by a factor $1/[1 + \lambda_k/R]$, where $R = 1 + r_f$ is the risk-free return, and $\lambda_k = (1 - q_k)\Delta_k$ measures the information diffusion about security (company) k , with q_k being the share of investors ‘knowing’ security k and Δ_k being the shadow cost (Lagrange multiplier) of ‘knowing’ security k . If all investors are informed of security k , then $q_k = 1, \forall k$, and so $V_k = V_k^*$. However, when there are investors who were not informed of that company, then $q_k < 1$ and the value of the company would be below the equilibrium with complete knowledge (the CAPM equilibrium). From there Merton deduces that the spread between the expected return on security k under incomplete (\bar{R}_k) and complete (\bar{R}_k^*) information would be proportionate to its shadow cost:

²⁷ Hence the market is complete.

$$\bar{R}_k - \bar{R}_k^* = \lambda_k (\bar{R}_k / R).$$

Aggregating across securities he arrives at the following Security Market Line (SML) equation under incomplete information

$$\bar{R}_k - R = \underbrace{(\lambda_k - \beta_k \lambda_M)}_{\alpha_k} + \beta_k (\bar{R}_M - R), \quad (3)$$

where R_M equals (one plus) the return on the market and λ_M is the weighted average shadow cost of incomplete information over all securities. Again, in case of complete information where $\lambda_k = 0$ for all k (and so $\lambda_M = 0$), the SML reduces to the well-known CAPM case. Hence the prediction of the model is that the less known securities should earn a spread over securities which were well known to the market.

The model comparative statistics (op. cit., p. 494-499) examines the cross-sectional differences implied by the model. In particular, if $\psi(Y)$ denotes the elasticity of the expected excess return for security k ($|\bar{R}_k - R|$) with respect to some parameter Y of the model, i.e. $\psi(Y) = d \log |\bar{R}_k - R| / d \log(Y)$, Merton (1987) deduces that $\psi(b_k) > 0$, $\psi(x_k) > 0$, $\psi(\sigma_k^2) > 0$, and $\psi(q_k) < 0$, where b_k is the loading of returns on security k on the common factor Y , x_k is the share of company k in the total market, σ_k^2 is the idiosyncratic variance of company k , and q_k is the share of investors 'knowing' security k .

Thus the model predicts that the expected excess returns are higher for companies of larger size (higher x_k) and higher exposure to the common factor (higher b_k). Since β_k is an increasing function of b_k ,²⁸ the latter implies that companies with higher beta should also earn higher expected return. On the other hand, more obscure companies (low q_k) should earn higher expected returns over the well-known ones. For our study, an important prediction of the model is that the positive elasticity of expected excess returns with respect of idiosyncratic variance (σ_k^2) means that idiosyncratic risk is priced and investors in companies with higher idiosyncratic risk earn a premium for assuming idiosyncratic risk. This relation is not without qualifications: the amount of the premium is not constant but dependent on how widely followed the security is. In equation (3) above the excess return depends on the magnitude of the shadow cost of knowing security k (λ_k) net of the average shadow cost of knowing the securities on the market (λ_M). Merton (1987) shows that $\lambda_k = (1 - q_k) \Delta_k$, where $\Delta_k = \bar{R}_k - R - b_k (\bar{R}_{n+1} - R)$ is the excess return on asset k relative to the predicted return from exposure on the market factor as captured in the forward asset dependent on the state of Y . For a given delta, the amount of the premium on holding security

²⁸ $\beta_k = (b_k b + x_k \sigma_k^2) / \text{var}(\tilde{R}_m)$, eq. 22 on p. 493 in Merton (1987)

k depends on the share of investors aware of the asset. In particular, we should not expect to find a constant premium for idiosyncratic risk in the cross-sectional regressions. If we restrict ourselves to widely followed securities, we should expect α_k close to zero. Thus, depending on the types of shares that we have in the test sample, failure to reject the null hypothesis that $H_0: \alpha_k = 0$ could in fact be a finding consistent with Merton's rather than a contradicting one.

Empirical tests of the proposition that idiosyncratic risk should be priced are hindered by the complex interplay between idiosyncratic risk and the rest of the variables in the estimation of equilibrium returns. Much of the empirical literature focuses on single point estimates of the idiosyncratic risk premium. Furthermore, the aggregation in portfolios using equal capitalisation of NYSE breakpoints could reduce the share of smaller, less known securities and could apparently reject the predictions of Levy (1978) and Merton (1987), whereas in fact the results could be interpreted as non-contradicting the two extensions of CAPM. In view of that, it seems particularly important to be cautious about discarding stocks from the sample, or aggregating in portfolios biased towards larger and presumably better known stocks, for which the idiosyncratic risk premium is predicted to be small or non-existent.

More recently, Malkiel and Xu (2004) revisited the problem of market equilibrium when investors are constrained from holding all securities. In their model the covariance matrix includes three (groups of) assets, hence the matrix has the form

$$V = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{pmatrix}.$$

There are three investors in the economy. The second investor could invest in all three assets (a , b , and c), while the first and the third investor are constrained from investing in assets a and c , respectively. Thus the variance matrices perceived by the constrained investors, V_{bc} and V_{ab} , are obtained from the original matrix V where the row and column corresponding to the prohibited assets are replaced by zeros. The covariance function perceived by the constrained investor then has the form

$$V_* = \left(\frac{n_1}{n_1+n_3} \begin{pmatrix} 0 & 0 \\ 0 & V_{bc}^{-1} \end{pmatrix} + \frac{n_3}{n_1+n_3} \begin{pmatrix} V_{ab}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1},$$

where n_1 and n_3 were the number of constrained investors in the first and the third group, respectively. All investors maximise quadratic utility function with risk-tolerance parameter τ , i.e. $u(W) = \mathbb{E}(W) - \frac{1}{2\tau} \text{var}(W)$. In that setting Malkiel and Xu (2004) demonstrate that

instead of the usual CAPM equilibrium $((\mu - r_f \mathbf{1}) = \frac{1}{n\tau} VS)$, where S is the supply vector for each security, the equilibrium relationship would instead have the form

$$(\mu - r_f \mathbf{1}) = \frac{1}{n\tau} VS + \frac{\eta_{1.3}}{n\tau} V\omega,$$

where $\eta_{1.3} = (n_1 + n_3)/n$ is the share of constrained investors, $n = n_1 + n_2 + n_3$ is the total number of investors, and $\omega = [(\mathbf{I} - V_*^{-1}V)^{-1} - \eta_{1.3}\mathbf{I}]^{-1}S$ is a supply adjustment. In particular, Xu and Malkiel (*ibid.*, eq. 13) show that under the assumption of uncorrelated idiosyncratic innovations, the idiosyncratic risk premium for security i has the form $(k\gamma\omega_{*,i}\sigma_i^2)$, where k is a constant, γ is the Sharpe ratio (market price of risk) that translates the idiosyncratic variance into risk premium,²⁹ σ_i^2 is the usual measure of idiosyncratic variance, and $\omega_{*,i}$ was the i -th element of ω vector. The presence of the product $(\omega_{*,i}\sigma_i^2)$ in the formula for the premium, rather than just the idiosyncratic variance, suggests that the risk premium for assuming idiosyncratic risk depends on the ‘supply adjustment’ for the particular security (they term this undiversified idiosyncratic risk), rather than the simple (total) idiosyncratic risk. As with the models of Levy and Merton, a principal problem in testing the model is estimating the price adjustment due to the unobservability of the reference portfolios against which investors measure risk and return. In addition to the cross-sectional tests of idiosyncratic risk, Malkiel and Xu (2004) propose to perform also time-series tests and to employ a work-around for testing the significance of idiosyncratic risk: to estimate the return on an idiosyncratic risk hedging portfolio. To that end they sort stocks into three size buckets and two volatility buckets in the spirit of Fama and French (1993) and Fama and French (1996) and find evidence for significant positive return on idiosyncratic risk portfolios.

A subtle implicit assumption of the CAPM and its extensions is the composition and observability of the market portfolio. That assumption was identified by Roll (1977), who pointed out that in any sample of assets there exist infinitely many mean-variance efficient portfolios, and the relation between returns and the betas with any of these portfolios would be linear, irrespective whether or not the market portfolio is mean-variance efficient. Moreover, the market portfolio must include not only the assets traded at stock markets, but also all other assets available to investors, including real estate, precious metals, commodities, and human capital. Building on that critique, Eiling (2013) suggested an extension of the CAPM pricing model to a multifactor pricing model that explicitly distinguished tradable and

²⁹ Normally the denominator of the Sharpe ratio is the standard deviation of the market portfolio; in this case, however, Xu and Malkiel defined the ratio as having the variance of the market portfolio in the denominator, i.e. $\gamma = (\mu_m - r_f)/\sigma_m^2$.

K non-tradable assets:

$$\mathbb{E}r_{tr,i} = \beta_{mkt,i} \left(\mathbb{E}r_{mkt} - \bar{\gamma} \sum_{k=1}^K Cov(r_{mkt}, R_{nt,k}) q_{nt,k} \right) + \sum_{k=1}^K \beta_{nt,k,i} (Var(R_{nt,k}) q_{nt,k}),$$

where $\mathbb{E}r_{tr,i}$ is the expected return on tradable asset i , $\bar{\gamma}$ is the market aggregate risk aversion coefficient, q_{nt} is the $K \times 1$ vector of aggregate wealth due to the nontradable assets divided by the total value of tradable assets, and $\beta_{mkt,i}$ and $\beta_{nt,k,i}$ are betas with the traded market and each of the K non-traded assets. This shows that if idiosyncratic risk correlates with some non-traded asset, then idiosyncratic risk could be priced by the market, but the reason for that premium is not idiosyncratic risk per se, but rather the correlation with the non-traded asset. In particular, Eiling (2013) finds that the non-tradable assets model with (orthogonalized) industry-specific human capital can reduce the premium for idiosyncratic risk by 36%.

In Table 1 we have summarised the principal theoretical models as they relate to this study. The predictions of the CAPM should be considered as a baseline case, upon which the remaining models extend. We have not covered the Arbitrage Pricing Theory developed by Ross (1976) because the theory as such does not identify the factors that drive the stock returns and thus would be of limited use for the present debate. Nonetheless, the structure of that model requires that the explanatory factors should not be diversifiable, and thus the predictions of that theory would overlap with those of the CAPM extensions, viz in absence of limits to diversification idiosyncratic risk would be diversified away.

Table 1: Summary of selected theoretical models and their predicted correlation between idiosyncratic risk and stock returns

Study	Predicted correlation	Comment
Sharpe (1964); Lintner (1965b)	None	The studies formulate the Capital Asset Pricing Model (CAPM) that predicts that in equilibrium idiosyncratic risk will be diversified away and stocks with higher idiosyncratic risk will not earn higher returns.

Study	Predicted correlation	Comment
Kahneman and Tversky (1979)	Unspecified	The authors develop the prospect theory of decision-making, which places higher value on extreme gains or losses rather than on expected lottery outcomes. Such a setting could allow higher returns on skewed outcomes, e.g. higher prices (lower returns) on securities with small probability of high gains, as in Barberis and Huang (2008).
Chamberlain (1983); Owen and Rabinovitch (1983)	Unspecified	The studies suggest that if returns are jointly elliptically distributed then investor preferences could be expressed as a function solely of expected return and variance, lending indirect support on the mean-variance optimisation approach of Markowitz (1952).
Levy (1978)	Positive	The model predicts that if transaction costs, taxes, or other market frictions prevent the full diversification of idiosyncratic risk, then in equilibrium the market should price idiosyncratic risk and stocks with higher idiosyncratic risk should earn, <i>ceteris paribus</i> , higher returns.
Merton (1987)	Positive	The model extends the results of Levy (1978) and derives comparative statics of the model, predicting the direction of correlation between idiosyncratic premium and security parameters (share of investors, size, beta with market, and share of investors following the security).
Malkiel and Xu (2004)	Positive	The study extends the results obtained by Levy (1978) and Merton (1987) and predicts that the premium on idiosyncratic risk would depend on the total idiosyncratic risk scaled with a supply adjustment.
Eiling (2013)	Unspecified / spurious	The author extends the pricing model of the CAPM with non-tradable assets, as suggested by Roll (1977) and finds that about a third of the premium to idiosyncratic risk is in fact due to exposure to industry-specific human capital.

Source: the author

In this section we explored the principal theoretical models that explore the pricing of idiosyncratic risk in frictionless and frictional markets. The development followed the historical evolution of the debate in order to reveal in context the implicit and explicit assumptions that underlie the CAPM. This enabled us to see the range of situations that could invalidate the key prediction of the model, namely – the irrelevance of idiosyncratic risk for the pricing of individual assets, namely: alternative decision-making rules than those assumed by von Neumann and Morgenstern (1944), e.g. the behavioural finance theories; inadequacy of the quadratic approximation to investor's utility function that could ascribe some role to higher moments like skewness and kurtosis; asset returns that are not jointly elliptically distributed so that investor utility function may not be specified in terms of expected return and variance alone. Special emphasis was placed on three studies that relate most directly with the topic of our enquiry – Levy (1978), Merton (1987), and Malkiel and Xu (2004). The three models offered alternative formulations of the portfolio-construction problem when investors are constrained from holding the market portfolio and from diversifying entirely their idiosyncratic risk. Those models provide some insight into how investor constraints and asset characteristics jointly determine the equilibrium prices and returns of the risky assets. Admittedly, unlike the CAPM which allowed analytical solutions, the idiosyncratic risk premia of those models depend on characteristics that we are unable to observe directly, like the limits on number of securities, the investment universe used by each investor in his portfolio formation, and ultimately – the true distribution of idiosyncratic shocks. Nevertheless, the models yield insight into the factors that could affect the correlation of the idiosyncratic risk premium with other asset characteristics. In the following section we shall review the findings of some key empirical studies on the correlation between idiosyncratic risk and stock returns.

2.3. Empirical findings concerning idiosyncratic volatility and the cross-section of stock returns

Empirical evidence on the composition of portfolios of investors diverges from the predictions of CAPM. Blume and Friend (1975) are among the first to document that contrary to the predictions of CAPM, investors hold rather concentrated portfolios. They examine a sample of 17,056 tax returns from the 1971 tax year, as well as the distribution of assets in the

Federal Reserve Board's 1962 Survey of the Financial Characteristics of Consumers, and find that many portfolios are highly undiversified. They report that as much as 34.1% of the tax returns list only one dividend-paying stock³⁰, and 50.9% list up to two dividend-paying shares. The results of the examination of the 1962 survey are similar: 50% of households holding at least 63% in one asset, excluding holdings in mutual funds; if investment in mutual funds is added as well, the concentration increases to 90% held in one asset.

Kelly (1995) studies portfolio diversification of a sample of 3665 households (the 'regular sample') from 1983 Survey of Consumer Finances, and another sample of 438 households from the top two percentiles of the income distribution. Of the 635 stockholders in the regular sample he finds that only 35 (i.e., 5.5%) hold 10 or more shares, and only 11 of these – 20 or more shares. In a regression analysis of the likelihood of holding more than 10 shares on a set of explanatory variables³¹ he finds that only the portfolio value and the number of trades are significant in all specifications, and the share of stocks in a household's financial assets is statistically significant only in some specifications.

In the high-income sample, Kelly (1995) finds that the median number of shares held (excluding shares in companies where household members works) is 10. With the increase of portfolio values, diversification improves and for the highest bucket in that study (\$1.3 million to \$52.5 million) the share of portfolios with over 20 stocks increases to 61%. Many more variables are found to be significant in the explanatory regressions, however the explanatory power of those regressions (as measured by the pseudo- R^2) is significantly lower.

Goetzmann and Kumar (2001) examine 79,995 equity investment accounts in a large (unnamed) discount brokerage house held by individual investors during a six-year period from 1991 to 1996. In order to exclude dormant accounts, the study focuses on 41,039 accounts that have at least five trades over the surveyed period. The aggregate value of the examined portfolios exceeds \$2.5 billion. The median portfolio in the database is \$13,869, and the average portfolio is worth \$35,629. The authors report that the average number of stocks in a portfolio is just 4, and the median number is 3. More than a quarter of the portfolios contain just 1 share, and less than 5% are reported to include 10 or more shares.

³⁰ For tax purposes households were required to disclose only the dividend-paying stockholdings; information on the ownership of non-dividend-paying shares was not available in that sample.

³¹ Portfolio value, number of trades, share of portfolio, age, college education, management job, self-employed, advice from broker, attitude to risk, term life insurance, whole life insurance, investment in mutual funds, defined contribution pension plan, IRA (as % of household financial assets), trust fund (as % of household financial assets)

The authors discuss whether those portfolios might be used for gambling on the stock market, but dismiss that as unlikely in view of the material balances on the accounts in absolute and in relative terms, compared to the annual household income. The authors find that the degree of diversification increases with the increase of age, possibly reflecting increasing risk aversion. They also find that the number of shares held increases with the value of the portfolio. However, investors are not found to use correlations when structuring their portfolios, which is found to result in suboptimal allocations. They report that just 15% of portfolio values are invested in mutual funds and do not find evidence that less diversified investors compensate the under-diversification by investing more in mutual funds. They conclude that the insufficient diversification observed in their sample could result in pricing of idiosyncratic risk, as suggested by the works of Goyal and Santa-Clara (2003) and Malkiel and Xu (2004).

Such findings are in stark contrast to the recommendations of various empirical analyses of required levels of diversification. Statman (1987) examines the diversification benefits of increasing the number of shares, assuming randomly constructed portfolios, and finds that a minimum of 30 stocks for borrowing investors and 40 stocks for lending investors are required for sufficient diversification.³² Other studies, however, suggest other thresholds. For example, an earlier study by Elton and Gruber (1977) indicates that the benefits of diversification are largely exhausted with a portfolio of just 10 to 15 securities. They observe that increasing the number of securities from 1 to 10 eliminates about 50% of variance, and increasing the number of assets to 20 improves that by just 5 percentage points.

Levy (1978) points out that the evidence on the composition of individual portfolios need not imply a rejection of CAPM itself, and that the model could be still valid on normative grounds as producing results that are consistent with the observed equilibrium expected returns. However, since the empirical evidence in favour of CAPM is mixed too, he examines the equilibrium expected returns when the number of assets, in which the investor could invest, was limited. Such a cap on the number of assets is consistent with the empirical evidence on the number of assets in portfolios, but could also reflect transaction costs or imperfect divisibility of assets.

Following the introduction of CAPM, a number of papers investigated whether the model held in practice, i.e. whether the expected market excess return was the sole factor

³² The numbers for borrowing and lending investors were different because the borrowing investors were assumed to borrow at the risk-free rate plus a marking of 2 percentage points while the lending investors were assumed to lend at the risk-free rate.

predicting asset returns. The first tests of the model were performed by Lintner (1965a)³³, who introduced a two-step procedure for testing the CAPM, which was further refined by subsequent studies. The first step uses cross-sectional regression of realised returns on an intercept and market returns: $R_{i,t} = a_i + b_i R_{m,t} + e_{i,t}$. At the second step, returns $R_{i,t}$ are averaged to obtain average asset returns \bar{R}_i , $\bar{R}_i = \frac{1}{T} \sum_t R_{i,t}$, and average returns are regressed on betas: $\bar{R}_i = a_1 + a_2 b_i + e_i$. Taking into account that CAPM is formulated in terms of excess returns, the expected values of a_1 and a_2 if CAPM holds are $a_1 = r_f$, $a_2 = R_m - r_f$. Lintner runs that two-step procedure using 301 stocks and returns for the period 1954-1963. In order to test if idiosyncratic risk is a priced by investors, at the second step Lintner (1965a) includes not only intercept and asset betas, but also the variance of the error term from the first-step regressions, $S_{e_i}^2$; if CAPM holds, one expects the variance term to be redundant and its coefficient to be insignificant. Lintner (1965a) obtains the following estimates:

$$\bar{R}_i = 0.108 + 0.063b + 0.237S_{e_i}^2$$

$$\begin{matrix} (0.009) & (0.035) \\ t = 6.9 & t = 7.8 \end{matrix}$$

These results provide partial support for CAPM. Indeed, they confirm that beta is a significant predictor of average asset returns, and the coefficient is positive, consistent with the theory. The slope of the regression equation (0.063) is significantly below the spread between market returns and the risk-free rate, which over the examined period is about 0.165. The intercept (0.108) is significantly higher than the risk-free rate. Finally, the coefficient of idiosyncratic risk turns out positive and statistically significant, contrary to the prediction of CAPM.

Miller and Scholes (1972) confirm the findings of Lintner (1965a) over an extended period. Furthermore, comparing the coefficient of multiple correlation they note that idiosyncratic variance alone has higher correlation with average returns compared to beta alone, with corresponding coefficients of multiple correlation of 0.28 and 0.19, respectively. When both beta and idiosyncratic variance are used in the second-stage regression, the coefficients of the two are positive and significant, and the coefficient of multiple correlation stands at 0.33. The estimated model, however, shows the same deficits as in Lintner's study – the value of the intercept is significantly higher than the risk-free rate, and the coefficient of beta is significantly below the spread between the market return and the risk-free rate:

³³ Reported on p. 192-95 in Levy (2012)

$$\bar{R}_i = 0.127 + 0.042b + 0.310S_{e_i}^2$$

$$\begin{matrix} (0.006) & (0.026) \\ t = 7.40 & t = 11.76 \end{matrix} .$$

More importantly, Miller and Scholes (1972) point out that the specification of the second-stage regression requires the use of the true betas but in fact uses the estimates from the first-stage regression, and therefore the betas are measured with some error and are in fact random. They show that this issue results in downward bias of the coefficient of beta. The first tests of CAPM that addresses this error-in-variable problem are performed by Black et al. (1972) and Fama and MacBeth (1973). Black et al. (1972) argue that correlations between the pricing errors of individual securities prevent the existing methodologies from testing the market model. To address these correlations, they propose to pool securities into portfolios of similar securities. Then the pricing errors for the individual assets would cancel one another, while any non-independence effect would be absorbed into the intercept; furthermore, the error in the estimation of betas should also diminish as a result of the aggregation, and the standard regression techniques and tests could be employed. In order to test the CAPM one needs observations for returns for various values of beta, hence Black et al. (1972) propose to form these portfolios based on values of beta, thus ensuring that the variance of beta will be the highest possible. They observe that creating portfolios using estimates of beta obtained from the tested sample would introduce sample selection bias; indeed, the portfolio containing the securities with highest betas would likely also be exposed to the highest positive pricing errors; similarly, the portfolio with the lowest betas would also be exposed to the securities with the lowest negative pricing errors. Therefore, they propose to use estimated beta from a previous time period as an instrument variable to construct portfolios, and then test portfolio returns using their subsequent returns, which were not used to estimate portfolio betas. They use monthly returns from the five years preceding the examined period to estimate betas that are then used for portfolio formation for the following one year. For example, for all securities that were traded as of January 1, 1931, and were traded for at least two of the preceding five years, they estimate security beta using the available history from January 1926 until December 1930. Then they form 10 portfolios based on the deciles of beta, and track the monthly returns for each portfolio for each of the next 12 months of 1931. Then the procedure is repeated with all securities traded as of January 1, 1932, with new betas estimated using data from 1927–1931, and formation of 10 new portfolios. In that way they obtain 35 years of monthly returns on ten portfolios from the 1,952 securities available to

them. For the second stage – the cross-sectional regression – they average returns over time periods and test the model with average returns on averaged betas.

Black et al. (1972) estimate the coefficients of the market model for each of the 10 decile portfolios for each of the 35 years, obtaining coefficients separately for each portfolio. They find that portfolio average return is indeed commensurate with portfolio beta. While most of the intercepts are insignificant, the authors warn that due to non-stationarity of the estimated coefficients, as well as some limitations of aggregations, the values of t -statistics might materially understate the significance of the results. Using the data for all portfolios they obtain the following pricing relationship:

$$\bar{R}_i = 0.00359 + 0.01080b$$

$$(0.00055) \quad (0.00052).$$

$$t = 6.52 \quad t = 6.53$$

The equation shows that market excess return is confirmed as a priced factor. The intercept (γ_0) in the above cross-sectional equation is significantly positive and different from zero, which contradicts CAPM. Black et al. (1972) hypothesise that this suggests the existence of a second factor. The time series regressions suggest that the intercept from time-series regressions decreases with beta, as noted by Black et al. (1972), and these results also hold, albeit in a weaker form, for the four sub-periods, except for the sub-period January, 1931 – September, 1939, when the relationship is inverted.

Fama and MacBeth (1973) further refined the method for testing CAPM, and their procedure became one of the most widely used methods for testing factor pricing models. The first step in their procedure is similar to the method used by Black et al. (1972). Fama and MacBeth (1973) firstly use seven years of data to form portfolios. Then for the next five years they re-estimate the time series regressions in order to re-evaluate the betas of individual securities. Finally, the last four years are used to fit the cross-sectional regressions. The end of the period formation and initial estimation windows are then recalculated, so that the next testing four-year period follows the preceding one without leaving gaps or intersecting it. In the second step they run a cross-sectional regression of returns on intercept, beta, squared beta, and idiosyncratic standard deviation (the standard deviation of the error term of the time-series regression). The last two terms are intended to test for non-linearity in beta and for the significance of idiosyncratic volatility as a pricing factor. Unlike Black et al. (1972), Fama and MacBeth (1973) estimate that regression separately for each month rather than averaging returns and betas in the testing period:

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t}\beta_{i,t-1} + \gamma_{2,t}\beta_{i,t-1}^2 + \gamma_{3,t}S_{i,t-1}.$$

The significance of the hypothesised factor loadings ($\hat{\gamma}_i, i = 1,2,3$) is tested using the standard t -statistic:

$$t(\hat{\gamma}_i) = \frac{\bar{\hat{\gamma}}_i}{s(\hat{\gamma}_i)/\sqrt{n}}$$

where n is the number of cross-sectional regressions, $s(\hat{\gamma}_i)$ is the standard error of the regression coefficient, and $\bar{\hat{\gamma}}_i$ is the mean value of the respective factors loading, averaged across all monthly cross-sectional regressions. Fama and MacBeth (1973) find that the factors for idiosyncratic volatility and the non-linearity term are not significantly different from zero. On the other hand, as pointed out by Levy (2012), the Fama and MacBeth (1973) methodology “employs portfolios rather than individual assets; therefore, it has the advantage of minimising the measurement errors in beta and the disadvantage of not testing asset pricing of individual assets. Thus, in the case of supporting the CAPM, one cannot generalise it to individual risky assets.” (p. 200)

Roll (1977) questions the previous tests of CAPM pointing out that the market portfolio to which CAPM refers is the portfolio of all assets in the economy, including human capital, real estate, privately held businesses, overseas assets. Clearly that portfolio is not observable by econometricians, and its replacement by equity index introduces an error-in-variable problem, which could account for observed anomalies detected by some of the empirical tests. Furthermore, within the limited number of assets traded at stock markets there is always a mean-variance efficient portfolio, and finding that a portfolio is or is not mean-variance efficient has no bearings on the validity of CAPM.

The recent resurgence of the interest in idiosyncratic risk is triggered by Campbell et al. (2000) who decompose individual stock returns into market, industry and idiosyncratic components under the assumption of constant betas equal to 1. Using data from the Center for Research on Security Prices (the CRSP data set) spanning the period from July 1962 to December 1997 they calculate the monthly volatility series constructed from daily data. They reject the unit root hypothesis at 5% confidence level for average market, industry, and idiosyncratic volatilities. They also find a positive and statistically significant trend in average idiosyncratic risk. Similar results are obtained using various frequencies (daily, weekly, monthly) and according to the authors are not attributable to outliers. The implications of the results of Campbell et al. (2000) are intriguing. They document that idiosyncratic risk increased from 65% of total risk to 72% of total risk over the period from 1962 to 1997. This increase suggests that correlations between individual assets should be declining, a phenomenon also documented by their study, implying increasing benefits from

diversification. Furthermore, the authors document cyclical volatility of all volatility components, with market, industry, and idiosyncratic volatilities all increasing during economic downturn.

Following the publication of Campbell et al. (2000), a number of authors published studies concerning idiosyncratic risk, raising the questions whether idiosyncratic risk is increasing over time, whether average idiosyncratic risk predicts market return, and whether idiosyncratic risk predicts the cross-section of returns, which is the question also pursued by this study.

In an important contribution to the debate Malkiel and Xu (2004) examine if idiosyncratic risk is a significant predictor of expected returns. In their study they find that in portfolio context (using the Fama–Macbeth methodology), CAPM beta is an important factor in explaining cross-sectional differences of returns, but its effect declines over time. They find that idiosyncratic risk is significant, irrespective of whether it is measured as volatility of residuals from CAPM regressions, or Fama–French three-factor regressions. They furthermore report that the size factor is dominated by the idiosyncratic risk as an explanatory variable, and while the two of them are significant predictors when used individually, when used jointly in multiple regression only the idiosyncratic risk variable remains significant (at 94% level). Running cross-sectional regressions with individual securities, they again find that idiosyncratic volatility is a significant predictor that outweighs the size factor.

Spiegel and Wang (2005) note that there is a significant negative correlation between liquidity and idiosyncratic volatility, and that expected stock returns are positively correlated with idiosyncratic risk and negatively correlated with liquidity. They estimate idiosyncratic risk using an Exponential Generalised Autoregressive Conditional Heteroscedasticity model (EGARCH) with the three Fama–French factors as predictors of excess returns using monthly data and employing two types of liquidity measures – “cost based” measures (Gibbs, Gamma, Amihud, and Amivest) and “reflective” measures like traded volume or number of investors holding a security. They find that when one controls for idiosyncratic risk, only the dollar-traded volume has some predictive power over the cross-section of stock returns, while idiosyncratic risk is consistently positively correlated to expected stock returns. Their results suggest that the impact of a one standard deviation change in idiosyncratic risk is on average between 2.5 and 8 times stronger than the impact of a corresponding one standard deviation difference in liquidity.

In an important contribution to the field, Ang et al. (2006) investigate whether idiosyncratic volatility explains the cross-section of market returns. They fit the Fama–French

three-factor model³⁴ to the daily excess returns on individual stocks and calculate idiosyncratic volatility as standard deviation of the error term over the preceding month. They use a ‘1/0/1’ portfolio formation strategy³⁵ in order to examine whether portfolios formed on total volatility and idiosyncratic volatility have significantly different yields. They find that stocks ranked on idiosyncratic volatility exhibit a consistently negative correlation with expected returns after controlling for various factors (size, book-to-market, leverage, liquidity, volume, turnover, bid-ask spreads, co-skewness, dispersion of analyst forecasts), across sub-samples (NYSE stocks only, NBER recessions and expansions, high and low volatility episodes, and sub-periods), and for various ‘L/M/N’ strategies. Their findings are puzzling as they contradict both CAPM (idiosyncratic volatility should not be priced at all) and Levy (1978) and Merton (1987) (idiosyncratic risk should earn positive risk premium).

In a follow-up contribution, Ang et al. (2009) provide further evidence from the G7 countries and other developed markets that the spread between the first and fifth quintiles of portfolio sorts based on idiosyncratic volatility is again negative, standing at -1.31 per cent per month after controlling for world market, and size and value factors.

The findings of the two studies of Ang et al. (2006, 2009) caused significant controversy and prompted many authors to re-examine their findings. Bali and Cakici (2008) investigate how idiosyncratic risk is priced using NYSE/AMEX/NADAQ data over the period July 1963 – December 2004 and find that the results of Ang et al. (2006) are not robust with respect to weighting scheme, time frequency, portfolio formation, and screening for size, price, and liquidity. Using equally-weighted instead of value-weighted portfolio returns, they find that the spread between the highest-risk portfolio and the lowest-risk portfolio averaged insignificant positive $+0.02$ percentage points. Similar small positive spread ($+0.08$) is also obtained using reciprocal idiosyncratic volatilities as weights for averaging. Furthermore, they point out that the capitalisation of the high-risk and low-risk quintile portfolios used by Ang et al. (2006) are quite different: the high-risk portfolio contains 20% of all stocks but accounts on average for just 2% of market capitalisation; in contrast, the low-risk portfolio comprising of 20% of the stocks with the lowest idiosyncratic risk accounts for 54% of market capitalisation. Using the NYSE quintile breakpoints results in more balanced portfolio

³⁴ The model is briefly discussed in the next chapter.

³⁵ The ‘L/M/N’ notation means that at moment t the idiosyncratic volatilities were estimated using ‘L’ months of daily data from month $(t - L - M)$ to month $(t - M)$; then at time t portfolios are formed based on the quantiles of the distribution of idiosyncratic volatility, and these portfolios were held for ‘N’ months.

capitalisations; with those NYSE breakpoints they find that all weighting schemes (value, equal, inverse of idiosyncratic risk) yield statistically insignificant spreads between high-risk and low-risk portfolios. The spread is also insignificant when the breakpoints are selected based on market capitalisations so that each portfolio has approximately equal market value. Bali and Cakici (2008) also observe that the results appear sensitive to calculation frequencies. For all breakpoints and weighting schemes, they find no significantly positive or negative average return spread for the NYSE/AMEX/NASDAQ stocks. Finally, they also argue that the results of Ang et al. (2006) are not robust with respect to price, size and liquidity factors. In the three sub-samples of large/liquid stocks, large/high priced stocks, liquid/high priced stocks they find no evidence for a significant link between idiosyncratic volatility and the cross section of expected returns. However, it should be noted that in Bali and Cakici (2008) idiosyncratic risk tends to lose its predictive significance when re-assigning stocks with high capitalisation, and with presumably higher visibility, to high-volatility portfolios. Such re-assignment should reduce the spread of yields between the portfolios, consistent with the comparative statics of Merton (1987), and hence make more difficult the rejection of the null hypothesis of no difference in smaller samples.

In an important contribution, Fu (2009) commented that the theoretically correct variable to explain expected returns is the expected idiosyncratic risk rather than the past idiosyncratic risk. He points out that in his sample the first-order autocorrelation of idiosyncratic volatilities is just 0.33, which according to him renders previous-month idiosyncratic volatility an unsatisfactory proxy for current-month idiosyncratic volatility. The same conclusion is reached when Fu (2009) tests the series of realised idiosyncratic volatilities for unit roots. The null hypothesis of unit root is rejected for 90% of the series in his sample³⁶, which implies that previous-month volatility is not an unbiased predictor of current-month volatility. In order to overcome the limitations of the approach of Ang et al. (2006), he proposes to use the Exponential Generalised Autoregressive Conditional Heteroscedasticity (EGARCH) model to forecast next-period expected idiosyncratic volatility. He addresses concerns raised by other authors that the pooling of securities in quantile

³⁶ Fu (2009) uses Dickey-Fuller test (see Dickey and Fuller (1979)) with drift using levels and logs, i.e.

$$IVOL_{t+1} - IVOL_t = \gamma_0 + \gamma_1 IVOL_t + \eta_t,$$

$$\ln IVOL_{t+1} - \ln IVOL_t = \gamma_0 + \gamma_1 \ln IVOL_t + \eta_t.$$

Fu required at least 30 volatilities to perform the test. The null hypothesis that $\gamma_1 = 1$ was rejected at 1% confidence level for 89.97% of the tested securities in levels ($IVOL_t$) and for 87.81% of the tested log-volatilities ($\ln IVOL_t$).

portfolios is one of the reasons of the puzzling results of the studies of Ang et al. (2006, 2009) by using Fama–French regressions with individual securities as assets. He finds that cross-sectional returns are significantly (in both statistical and economic sense) positively correlated with idiosyncratic returns, with stocks with idiosyncratic volatility of one standard deviation higher than the average, earning a risk premium of approximately 1 percentage point, and a zero-investment portfolio that is long in the 10% of the stocks with the highest idiosyncratic volatility and short in the 10% with the lowest idiosyncratic volatility earns on average a premium of 1.75 percentage points per month. He argues that reversals of idiosyncratic volatilities are a significant contributing factor for the puzzling results obtained by Ang et al. (2006). The findings of Fu (2009) are qualitatively similar to the results of Bali et al. (2011) who report that once they control for the maximum positive return in the previous month, the anomalous regression slopes reported by Ang et al. (2006) are reversed.

One limitation of the study of Fu (2009) that is identified by Huang, Liu, Rhee and Zhang (2012) is the omission of lagged return as an explanatory variable in the cross-sectional regression, which could result in omitted-variable bias. Furthermore, the study design did not allow for explicit analysis of how idiosyncratic risk premium changes with the profile of investors holding the security. In a follow-up study, Fu and Schutte (2010) revisit these issues. Idiosyncratic risk is confirmed as a significant factor of the cross-section even in the presence of the lagged return. Consistent with the underlying economic theory, the slope for idiosyncratic risk in the cross-sectional regressions is found to be higher in samples involving securities with higher individual ownership and small size of orders (an indication of individual investing). On the other hand, for stocks with significant institutional ownership there is no robust link between idiosyncratic risk and returns.

Brockman et al. (2009) extend the study of Fu (2009) by examining a set of 44 international markets including 58,000 shares and spanning 27 years. Using Fama–Macbeth cross-sectional regressions and individual securities as assets, they find that idiosyncratic risk (measured in terms of a monthly EGARCH model fitted on the full series rather than on expanding windows), correlates positively with expected returns. Furthermore, they estimate the link between the idiosyncratic risk premium and various explanatory variables like turnover, breadth of analysts' forecasts, trading costs, errors of earnings forecasts.

The use EGARCH models to predict volatilities in the methodology of Fu (2009) is questioned by Guo et al. (2014); they argue that Fu uses month- t data to estimate the parameters of EGARCH. This in their view introduces a look-ahead bias, which may have

impact on the observations with higher return (as the likelihood of significant deviation is small, it implies material change in the parameters of the EGARCH model). Once this look-ahead bias is controlled, Guo et al. (2014) find no significant relation between idiosyncratic risk and the cross-section of returns.³⁷ The authors also point out that their criticism also extends to the use of EGARCH with full sample.

Brooks et al. (2011) examine how the benefit of diversification increases with the number of securities in an investor's portfolio. They construct portfolios with a number of securities ranging from 15 to 70 with an increment of 5 and run Fama–Macbeth cross-sectional regressions (with only beta and idiosyncratic risk as dependent variables). They find that idiosyncratic risk is priced in portfolios of up to 50 securities; after that threshold idiosyncratic risk becomes insignificant. However, one should note that these cross-sectional tests seem to confirm that the idiosyncratic risk premium is concentrated in a relatively small number of smaller securities with limited investor base and that construction of portfolios with a large number of securities blurs the risk-return relationship along the lines suggested by Ang et al. (2010). A better understanding of the implications of under-diversification on equilibrium prices might be achievable via simulations. Relative to the three-factor model of Fama and French (1993), the study finds that idiosyncratic risk pricing is detectable only for portfolios containing less than 20 shares. Overall, these results may suggest that using a constant number of portfolios in studies of the cross-section would result in an increasing number of assets in each cell and might result in loss of explanatory power of some variables over time.

Eiling (2013) explores one possible explanation for the significance of idiosyncratic volatility. She builds on the observation of Roll (1977) that the true market portfolio in CAPM includes all available assets, and in particular – human capital. She develops a model that suggests that systematic risk related to industry-specific human capital is omitted as a source of risk and ends up in the idiosyncratic residuals. She suggests that this mechanism might explain the predictive performance of idiosyncratic risk in the cross-section of returns. She calculates the monthly human capital returns in terms of per-worker labour income growth using two-month averages. She finds that industry-specific human capital explains between 10% and 36% of the idiosyncratic risk premium.

Cao (2010) and Cao and Xu (2010) argue that a distinction should be made between

³⁷ We shall deal with this criticism and how our approach avoids this look-ahead bias in the methodology section.

long-term and short-term risk. They suggest that it is the long-term idiosyncratic risk that actually explains the cross-section of returns. However, the interpretation of their results is obscured by the use of a Hodrick–Prescot filter to extract the cyclical component of volatility using parameters applicable to quarterly macroeconomic series. Furthermore, the long-term volatilities from the Hodrick-Prescot filter are obtained using the full series and are thus exposed to look-ahead bias (cf. the critique of Huang et al on the approach of Fu above), with different horizons of application of the filter resulting in a different split of the overlapping volatilities into long-term and short-term components.³⁸ More importantly, the outcome of the application of an HP filter may be very closely related to the residuals from the simple OLS regression. Indeed, observe that squared idiosyncratic returns are an estimator of idiosyncratic volatility, and therefore the variance of the residuals from the OLS regression could be interpreted as a moving average filter of the trailing series of volatilities. In the absence of a strong trend in volatilities³⁹, the output of the HP filter would likely be close to the moving average, and thus while the arguments of Cao (2010) and Cao and Xu (2010) are valuable, their evidence cannot be seen as conclusive.

A related study by Ruan et al. (2010) explores the idea that the contradictory empirical results are due to noise in the estimated volatilities by applying a dual predictor approach to estimating the unobservable aggregate idiosyncratic risk from the volatilities of equally-weighted and value-weighted portfolios of different sizes. In that setting the authors find that aggregate idiosyncratic risk is a significant predictor of market excess returns.⁴⁰

Huang, Liu, Rhee and Zhang (2012) base their forecasts on an ARIMA model fitted using realised monthly volatilities calculated by the method employed by Ang et al. (2006). They test the hypothesis with four estimators of idiosyncratic volatility and find that

³⁸ To clarify this point, consider a security X over the periods 2000-2008 and 2000-2016. If we use the HP filter over the longer period (2000-2016), the resulting split of volatilities into short-term and long-term components over the sub-period 2000-2008 would be different from what would be obtained using data only from the period 2000-2008. Furthermore, the values from the sub-period 2009-2016 would have affected the split in the first sub-period, 2000-2008, potentially introducing look-ahead bias. Even when the method is applied on expanding window designs, the current split into long-term trend and noise or cyclical component depends on the future.

³⁹ If there was a significant trend, this would have manifested in a difference in the Augmented Dickey Fuller tests with and without trend, but our analysis did not indicate such a problem.

⁴⁰ Note that this study explores a related but still distinctly different problem from ours, i.e. whether aggregate volatility predicts aggregate returns, which was pioneered by Goyal and Santa-Clara (2003)

introducing the lagged return ($R_{i,t-1}$) as a control variable, omitted in other studies, renders idiosyncratic volatility (lagged or predicted by the ARIMA model) statistically insignificant. In contrast, they report a significant positive correlation between the predicted volatility from the EGARCH model and the expected return.

Li et al. (2014) investigate whether there are significant trading gains from holding a portfolio that is long in low-risk securities and short in high-risk securities. They find that the composition of the high-risk portfolio changes quickly and the associated trading costs, as well as the small capitalisation and low liquidity of the high-risk stocks, wipes out most, if not all, gains from trading strategies aimed at exploiting the low-volatility anomaly.⁴¹ Their portfolios are formed based on the estimator of Ang et al. (2006), and the criticism that there are significant return reversals of high-volatility stocks cast doubt on the empirical significance of their results.

In a recent contribution, Fan et al. (2015) explore whether idiosyncratic volatilities are related to a range of asset pricing anomalies, including asset growth⁴², book-to-market value⁴³, investment-to-assets⁴⁴, short-term return momentum⁴⁵, new stock issues⁴⁶, size⁴⁷, and total accruals effects⁴⁸. Using five-by-five quintile portfolios constructed on idiosyncratic volatility and each of the anomaly dimensions, the study finds a significant link between idiosyncratic risk and the analysed stock anomalies, and that the impact of idiosyncratic risk on stock anomalies in developed countries is significantly weaker than its impact in developing markets.

⁴¹ In their view it was the securities with low volatility that earn abnormally low return, hence the reference to low-volatility anomaly.

⁴² The anomaly concerns the finding that companies experiencing higher asset growth earned lower returns; see Cooper et al. (2008)

⁴³ The book-to-market value anomaly concerns the finding that companies with high book/market ratio earn higher returns than companies with low book/market ratio; see Fama and French (1992), 1993), 2007)

⁴⁴ The anomaly concerns the finding that companies with a high investment-to-assets ratio earn lower returns than ones with a low investment-to-asset ratio; see Lyandres et al. (2008)

⁴⁵ The anomaly concerns the finding that securities with higher cumulative past returns (e.g. over six months) were likely to continue to perform better; see Jegadeesh and Titman (1993)

⁴⁶ Concerns the finding that companies with newly issued stocks underperformed otherwise comparable securities; see Fama and French (2008)

⁴⁷ Concerns the finding that companies with smaller market capitalization earned higher returns than those with larger market capitalization; see Fama and French (1992), 1993), 2007)

⁴⁸ Concerns the finding that cash flows and accruals are not fully reflected in prices until they impact prices, and thus companies with high total accruals earning lower returns than companies with lower total accruals; see Sloan (1996)

Table 2: Summary of selected empirical studies of the link between idiosyncratic volatility and stock returns

Study	Correlation	Comment
Fama and MacBeth (1973)	None	The study introduces the cross-sectional methodology for asset pricing tests. Using portfolios as assets, no significant correlation between idiosyncratic volatility and expected returns is found
Malkiel and Xu (2004)	Positive	The authors propose an extension of the Merton model and using monthly data on individual US equities and a sample of Japanese equities find positive correlation between undiversified idiosyncratic risk and expected returns. Idiosyncratic risk is reported to be more reliable a predictor than beta or size
Ang et al. (2006, 2009)	Negative	The study uses portfolios formed on daily idiosyncratic volatility in the preceding month. The authors find that for US & G7 countries, high-volatility securities have abnormally low expected returns compared to low-volatility equities. They report strong co-movement of negative returns for high-volatility stocks.
Fu (2009); Fu and Schutte (2010)	Positive	The studies use the EGARCH model to estimate expected returns. The papers use the methodology of Fama and French (1992) with monthly data and individual securities as assets and find robust positive correlation between expected idiosyncratic volatility and realised returns. The study of Fu also examines the stationarity of the volatility series and finds that the unit root null hypothesis is rejected for 90% of all securities. The study attributes the negative spread obtained by the studies of Ang et al. (2006) to return reversals.
Spiegel and Wang (2005)	Positive	The paper uses EGARCH model to forecast volatility and documents negative correlation between liquidity and idiosyncratic volatility. However, the impact of idiosyncratic volatility is 2.5 to 8 times stronger than the impact of the liquidity characteristic.

Study	Correlation	Comment
Bali and Cakici (2008)	Mixed	The study finds that data frequency, breakpoints, weighting schemes, and data filters all affect the significance and sign of correlation and concludes that the evidence is mixed and not convincing.
Cao (2010); Cao and Xu (2010)	Positive with the long-term component.	The dissertation and the working paper propose that long-term volatility as estimated from a digital filter explains the cross-section of returns, while the link between short-term volatility and returns is negative. The estimated split is, however, not forward-looking and is exposed by construction to look-ahead bias.
Li et al. (2014)	Economically insignificant	The authors find that the gains obtainable from holding low-risk minus high-risk portfolios are small and likely to be wiped out by frequent rebalancing and trading costs for small, low-liquidity securities.

Source: the author

2.4. Conclusions

In this chapter we covered two complementary aspects of our study. Firstly, we surveyed the theoretical models concerning our research problem. We found that these fall into two related strands: firstly, the Capital Asset Pricing Model (CAPM), which predicted that in a frictionless market investors solving the mean-variance portfolio optimisation problem would invest in the market portfolio, and that individual equilibrium returns would be determined solely by the beta of the asset with the market, and not by idiosyncratic risk, which investors can diversify completely. Secondly, when transaction costs, asset indivisibility or other frictions prevent investors from completely diversifying away idiosyncratic risk, idiosyncratic risk should be priced by the markets and the lower the number of assets in investors’ portfolios, the higher the equilibrium return. These models are refinements of the CAPM resulting from relaxation of its assumption that markets are frictionless.

The theoretical models alone are unable to resolve the problem whether idiosyncratic

risk is priced by markets. There are a couple of reasons for that: firstly, there is a trade-off between model tractability and realistic assumptions, e.g. assuming a two-period economy with deterministic volatility that is known to all investors; secondly, the premium for idiosyncratic risk is a non-linear function of the limits on diversification imposed by market frictions; thirdly, idiosyncratic risk is unobservable. Thus, we see that while the theoretical models provide a solid framework for analysis of equilibrium prices and returns, the problem of the correlation between idiosyncratic risk and returns is ultimately an empirical one.

The survey of the empirical evidence concerning the predictability of the cross-section of stock returns by idiosyncratic risk reveals that although significant contributions have been made towards understanding the patterns of market returns. Nonetheless, the existing empirical evidence is mixed, with some studies finding no robust correlation between idiosyncratic risk and returns, other reporting economically and statistically significant positive correlation, yet other studies finding a negative correlation. Therefore, some important questions concerning the link between idiosyncratic risk and returns remain unanswered.

The first of these questions is how the various idiosyncratic volatility estimators compare to one another. Indeed, few studies employ comparable estimators of idiosyncratic volatility, which invites the question how these different measures correlate with each other and with the unobservable true volatility. The answer to that question would elucidate whether the explanatory power of the different estimators lies in their correlation with the expected volatility or with other characteristics of the volatility dynamics. The second question is how the recent critiques on previous studies on methodological grounds (e.g. omitted-variable and look-ahead biases) impact the findings of (in)significance of idiosyncratic volatility in explaining the cross-section of returns.

In the following chapter we shall formulate our methodology for answering those empirical questions and we shall elaborate on how it relates to existing studies.

3. Research Methodology and Data Sources

3.1. Introduction

“I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.”

Lord Kelvin, 1883

The preceding chapter suggested that both nil and positive correlation between returns and idiosyncratic risk would be compatible with underlying economic models. However, empirical evidence thus far did not provide unequivocal guidance on the direction or strength of that correlation. We saw that while existing studies employed the same data set (the series of the Center for Research in Security Prices, augmented with Compustat data), the studies diverged greatly in terms of how to measure idiosyncratic risk and how to test the propositions of the underlying theoretical models. In this chapter we develop the methodological aspects of our empirical examination. We start with Section 3.2 that discusses and motivates the choice of our research methodology, which can be described as positivist, deductive statistical study. In that section we also highlight the strength and sacrifices that this choice imposed upon our study. Section 3.3 then discusses the choice of factor model and its estimation, as well as the selection of control variables for the cross-sectional regressions. Section 3.4. elaborates on the sources of data, as well as the calculation procedures for the calculated covariates. Section 3.5. outlines the Markov chain methodology for classification of volatility regimes, and provides an overview of the volatility episodes in the surveyed period. Section 3.6 concludes the chapter.

3.2. Methodological notes

3.2.1. Research philosophy: positivism and its limitations

The rise of positivism in economic research dates from at least the early 20th century. Inspired by the success of natural sciences, positivist ideas became increasingly adopted in economics, which eventually lead to the separation of economics from its precursor – the political economy. Economic research became increasingly concerned with the search for objective laws governing the economy, which were not particular to a given economic or political system, but prevailed universally in every economic system, irrespective of space, time, and politics.

Asset pricing theory, as a branch of finance theory, followed a similar pattern of development. Ryan et al. (2002) note that most of finance studies adopt a capital market perspective, which is reflected in the treatment of transactions and agents and investments and investors, while relatively few studies adopted a managerial or intercompany perspective (p. 51-52). They highlight three core propositions of finance: firstly, individual economic agents are formally rational; secondly, financial markets are perfectly competitive; and thirdly, information is freely available to agents. (*ibid*, p. 51)

The positivist approach to analysis of the social phenomena, however, is open to certain criticisms pointing out that social reality is constructed by humans and cannot be modelled with the deterministic certainty characteristic of most of physics. Salmon et al. (1999) highlight three alternative approaches to positivism in the context of social research: “Contemporary critics of the view that studies of human behavior are or can be scientific fall into three categories. The first group, called *interpretivists*, claim that explanations of human behavior are structured entirely differently from explanations of the behavior of physical objects since human behavior, they say, consists of actions done for reasons rather than events resulting from causes. The second group, called here *nomological skeptics*, do not deny that human behavior is subject to causal laws, but doubt that it will ever be possible to find laws of human behavior that are similar in power and scope to those in physical science. The third group, called *critical theorists*, claim that it is inappropriate even to try to explain human behavior in terms of the laws of cause and effect because to do so denies the value of human autonomy (free will). In addition, they say, any attempt to construct a social science on the model of physical sciences promotes unethical manipulation of humans and discourages any attempts to improve the conditions of social life.” (p. 407-408)

The positivist approach to economics (and finance) is not free of criticisms. For example, Caldwell (1980) outlines three criticisms to positivist economics. His first critique concerns the confirmability of such propositions. He points out that a rigorous test of a positive deductive argument requires clear specification of all assumptions and auxiliary hypotheses, but that is difficult to achieve, especially in social sciences, as there might be many control variables and assumptions. Secondly, Caldwell (1980) points out the distinction between theoretical terms in the positivist axiomatic-deductive models and the real world. Theoretical models work with theoretical terms, which may not exist in the real world, or may be unobservable. In our case the CAPM and Merton (1987) models work with abstract, theoretical concepts like the return-generating process, idiosyncratic risk, conditional homogeneous beliefs, two-period economy, or complete markets. Some of these are convenient assumptions that allow us to solve the model and produce testable predictions, while others are theoretical constructs with no direct match in the real world. Therefore, even in the core term of this study – idiosyncratic risk – we need to find a way to operationalize the term and find some way to measure it. However, there is no unique or correct way to do that – all studies in the field employ equally valid measures of idiosyncratic risk, yet they reach starkly different conclusions. This requires understanding how these alternatives relate to one another and to the underlying model, in order to interpret the results. Nevertheless, such interpretation is necessarily subjective and could be rejected by other researches, leading to different conclusions regarding the validity of the underlying positivist model. Finally, Caldwell (1980) notes that the role of scientific explanations may not fit nicely into the two models – the deductive-nomological or the inductive-probabilistic explanations.⁴⁹

We recognise the validity of the criticisms of positivism reviewed by Salmon et al. (1999) and Caldwell (1980), and we concur that economic ‘laws’ are qualitatively different from those of classical mechanics. Nevertheless, it should also be recognised that a number of results in economics can be found to be statistically valid, and although the next state of the social system cannot be predicted in the mechanical sense, some states are more likely than other. In a context akin to ours – a study of evidence in favour of technical analysis – Aronson (2007) points out that scientifically valid predictions should pass the discernable-difference test, so that they are either true or false, and the scientist should be able to discern between the predicted outcomes. Such a test requires two ingredients: the ability to reproduce objectively the assertion and the ability to distinguish between the ex-post events. In particular, theories

⁴⁹ For a more recent review of the explanatory schemes, see Woodward (2014)

that rely on subjective interpretation of information (e.g. stock prices) are not reproducible and may be consistent with conflicting outcomes; such theories might be true but would not be scientifically testable.

In some studies such a situation may be difficult to avoid, but in the present study the predicted outcomes (prices and returns) are readily observable, which might suggest that positivist philosophy might be a more useful context for our exploration. However, adoption of the positivist approach would also require that not only the outcome (returns) be objective, but also the risks that affect returns. One approach could be interpretivist and seek to elucidate how investors perceive risk and how they assess the risk of their holdings. Such assessment is likely to be a multicriterial one and could include how prices change in normal conditions (e.g. stock volatility), under stress (e.g. tail losses), soft facts about the company and its management, market strategy of the management team, transparency and reliability of reporting, and many more. Such an approach could uncover how investors assess risk much more reliably and realistically than simple standard deviation of returns. On the other hand, it would be difficult to extend those findings on the link between risk and returns, because that link is a statistical one and its validation would require a large sample of both high-risk and low-risk securities.

Summing up, we recognise that in the present context the positivist analysis has its limitations. These relate to its objective to ensure that the uncovered principals are universally valid across issuers and in time. Therefore, in our study we aim to ensure maximal size of our studies sample. Furthermore, we need to ensure that all our conclusions are reproducible, so we cannot rely on soft facts that might have different interpretation by other researchers. Despite these limitations, we shall nevertheless employ the positivist approach in order to pursue the external validity of our conclusions and to facilitate comparison with the existing body of literature surveyed in Chapter 2. This necessitates that we limit the basis of risk assessment in this study to objectively reproducible measures of risk that can be estimated for significant part of the market, and hence is based only on public-domain information that was available to all market participants (e.g. stock prices or publicly available financial reports). Moreover, we will focus on single measure of risk – the idiosyncratic volatility - and we shall not explore how investors ‘weigh’ the different measures in order to reach their overall risk assessment of the available securities.

3.2.2. Deductive and inductive research

The scientific arguments are broadly classified into two streams, depending on their approach: inductive and deductive ones. Inductive reasoning starts with analysis of individual instances and seeks to generalise the observations from individual instances into fundamental principles or regularities. Deductive reasoning, on the other hand, starts with some fundamental law and seeks to operationalise or confirm it for some specific instances.

Inductive reasoning starts with the observations of the world and notes patterns in it. These observations are then combined to build a theory of the explored phenomenon. This theory can be confirmed and refined by exploring its implications in other settings and examining whether the predicted outcomes are consistent with the actual data.

Deductive reasoning builds knowledge in the opposite way. It builds a theory of the world that produces some falsifiable predictions. Then the researcher performs tests in order to test if the general theory can be rejected. An example of that approach could be the development of the Expected Utility Theory, which started with a set of axioms concerning the behaviour of rational decision-makers under uncertainty, and produced a set of predictions that could be tested, as well as deduced implications from those axioms – the possibility to represent preferences in terms of utility functions.⁵⁰

Both inductive and deductive reasoning have their strengths and weaknesses. The principal characteristics of the inductive approach, as summarised by Salmon et al. (1999), are that it is ampliative⁵¹, but not necessarily truth-preserving⁵² and erosion-proof⁵³, and its arguments come with varying degrees of strength (p. 11). On the other hand, deductive reasoning has the advantages of being truth-preserving and erosion-proof, but non-ampliative and absolute in the sense that the deduction is either completely valid, or completely invalid.

Aronson (2007) notes that the most common type of inductive reasoning is induction by enumeration, which enumerates the evidence from some set of data and generalises it to some principle. For example, we may observe that in a sample of securities, those with higher idiosyncratic risk earned on average higher return, and therefore we could conclude that securities with higher idiosyncratic risk earn higher returns. Such generalisations should be made with care in order to ensure that they are not based on a non-representative sample (e.g.

⁵⁰ See Chapter 3, pp 57–62 in Ryan et al. (2002) for an exposition of the methodological tradition underpinning the modern theories of asset pricing.

⁵¹ i.e., the conclusions of the inductive reasoning go beyond the content of its premises

⁵² i.e., correct premises may lead us to wrong conclusions

⁵³ i.e., new premises may completely undermine a valid inductive conclusion

encompassing only periods of economic growth, or one including only large companies). Furthermore, the analysis should also take into account other characteristics and considerations that may have affected the observations, e.g. liquidity. Aronson (2007) points out that the most common error in inductive research is the hasty generalisation due to a small sample or low quality of evidence.

A problem with inductive financial market research is the data mining bias (or data snooping). Data mining is the name of a group of techniques that aim to uncover patterns in large data sets. Those methods have extensive applications in machine learning and artificial intelligence studies. However, such algorithms may also “overfit” the data, e.g. by using a too complex neural network; the outcome would be that the in-sample performance could be good, but that performance would not be sustained out of sample.

Due to the standardisation of stock market contracts and the available public information on prices, it has grown to be probably the most scrutinised market. Nevertheless, complete information is available for only a limited number of years and stocks. As a result, Black (1993) warned that the growing research in finance using the common data set (the data of the Center for Research in Security Price and the Compustat data on financial ratios) is likely to uncover patterns in the data where none exist. The ‘data mining bias’ therefore refers to the risk of finding spurious patterns in data that do not generalise out of sample. The sample-splitting approach used in the field of machine learning could be applied in finance as well, in order to mitigate that problem. However, McLean and Pontiff (2016) find that the abnormal returns to many of the stock market “anomalies” discovered by academic research shrink rapidly after publication, although they do not disappear entirely. This suggests that either the anomalies were due to data snooping⁵⁴, or patterns lose their predictive powers as investors attempt to exploit the anomaly.

The latter explanation of the shrinking of market anomalies after publications is particularly important in terms of choosing between inductive and deductive research. If investors change their investment behaviour over time, then the results of many techniques used in inductive research might not generalise well in time.

Deductive research starts from a theory explaining some phenomenon and operationalises it to specific instances in order to explain a certain phenomenon or validate the

⁵⁴ Researchers in finance rarely employ splits of the full sample into training, test, and validation subsets. Instead, they control for spurious results through robustness check using subsamples of the original sample in various dimensions or through reporting data for sub-periods.

theory. Deductive research is less concerned with social context compared to inductive studies and does not create new knowledge (non-ampliative), but its results are, in principle, erosion-proof and therefore it generalises well. It is the approach that we select for this study, which is motivated by the fact that it generalises and it is less exposed to data mining problem. Inductive research is more sensitive to social context and thus able to accommodate soft facts like attitudes and perceptions, but in view of the evidence of McLean and Pontiff (2016) it would be difficult to generalise the findings both in the past and in the future. Likewise, the small sizes implied by some research methods (e.g. interviews and questionnaires) mean that the conclusions may not generalize over the entire markets over the studied decades, and therefore the study would not be able to convincingly resolve the puzzling conflicting evidence on the correlation between idiosyncratic risk and returns. On the other hand, inductive use of techniques like regressions or neural networks could result in data mining, which may uncover patterns in the past, but those patterns could perish in the future, as documented by McLean and Pontiff (2016), unless there is some fundamental reason underpinning the pattern.

Another benefit of deductive research in that setting is the ability to make generalisations about the future developments on the market. For example, in recent years we observe various trends in the market that could affect the magnitude of the risk premium accruing to investors. The advances in information technology result in gradual reduction of transaction costs, which should encourage broader diversification and hence reduce undiversified idiosyncratic risk. On the back of automated trade execution, the recent years see increasing use of trading algorithms that shorten investment horizons and could be more able to trade on volatility changes. Another recent trend closely intertwined with the former two (reduction of trading costs and algorithmic trading) is robo-advising, which could increase the share of investment universe followed by smaller investors and promote their diversification, thus reducing idiosyncratic risk premium. Thus, the guidance of the theoretical model allows not only an exploration of the *status quo*, but also anticipates how technological and institutional advances on the marketplace could impact idiosyncratic risk premia.

We demonstrate the deductive research workflow using the present study as an example in Table 3. The table shows the key decisions that need to be made at each phase of the research: choosing a theory that could explain the link between idiosyncratic risk and returns; formulating the research hypotheses; operationalizing the theory; performing the tests; drawing conclusions and refining the theory.

Following the context phase, which in our case was concerned mostly with literature research and fact-finding, the first important decision in the deductive research workflow was choosing a theoretical model. As a baseline case we considered the Capital Asset Pricing Model (CAPM), which made a positive prediction of what determines stock returns. The CAPM extensions of Levy (1978) and Merton (1987) incorporate the assumption of incomplete diversification, while incorporating CAPM as a particular case. Thus, they appear as a natural choice for a theoretical model. Nevertheless, it should be recognised that there were other possibilities. Thus, Iwasawa and Uchiyama (2013) explore a behavioural finance explanation of the volatility anomaly in the Japanese market. Their behavioural exploration is based on the noise trader approach to finance in De Long et al. (1990), which proposes that a deviation of market prices from fundamental values may persist in markets where there are traders who trade without reference to the fundamental value of the traded assets ('noise traders') and the capacity or willingness of the arbitrageurs to take sufficiently large risky bets to close the gap between current price and fundamental value is limited.⁵⁵ In that setting Iwasawa and Uchiyama (2013) note that "[a]pplying this logic to interpret the 'volatility anomaly,' we can view that investors who like to buy stocks with high beta and/or high idiosyncratic volatility, even if they are overvalued, cause overpricing of these stocks on average, and low long-run average return of them." (p. 465) That market prices are more volatile than price fundamentals is a known fact, and the noise trader approach could accommodate such stylized facts. However, such an explanation is too ad hoc and non-specific, in the sense that a hypothetical proposition that investors like big companies or ones with stable cash flows and are buying them even when they are overvalued could be equally consistent with the noise trader approach. Therefore, we recognise that behavioural finance is motivated by actual behaviours and imperfections observed on the markets, although it is still not sufficiently mature to match the rigour of the traditional approach. The debate between the proponents of the two approaches goes well outside the scope of this discussion⁵⁶ and we are willing to agree that in cases where a theoretical model is not

⁵⁵ E.g. because trading involves assuming risk and the market price could go against the traders, causing significant loss to the arbitrageurs.

⁵⁶ The following quote from Subrahmanyam (2007) offers a fair summary of the points of the two sides: "Traditional finance academics often offer a few common objections to behavioural finance. First, it is often said that theoretical behavioural models are somewhat ad hoc and designed to explain specific stylised facts. The response is that behavioural models are based on how people actually behave based on extensive experimental evidence, and explain evidence better than traditional ones. Another common objection is that the empirical

consistent with empirical evidence, a new or refined model should be developed. However, we also believe that valid scientific models should produce testable non-trivial (non-obvious) predictions and in this instance the present author is more convinced by the arguments of classical finance. The concept of noise-trading may be consistent with some behaviours observed in the markets, but the limits to arbitrage do not seem consistent with the evidence of McLean and Pontiff (2016) of a significant decrease of anomalous gains after publication. Another relevant prediction could be based on the theory of Barberis and Huang (2008) who build on the cumulative prospect theory and argue that in the presence of demand for lottery-like stocks (i.e. ones with significant positive skewness that have large growth potential) gives rise to the own skewness of the stocks being priced and stocks with high positive skew being overpriced. Nevertheless, from the perspective of traditional finance it is not immediately clear what is the non-trivial insight from that theory, in the sense that if a model assumes preference for a certain type of assets, it will predict that that preference would be priced. Therefore, evaluating the arguments of both sides we prefer to base our deductive research on classical finance model, while recognising that behavioural models could also offer an alternative view on the matter.

The next step in the deductive workflow was the formulation of a research hypothesis. It was treated in the previous chapter and we shall not reproduce it here, but in summary we shall only note that since the theoretical model predicts that idiosyncratic risk is priced in the presence of under-diversification, this was naturally our principal research hypothesis. Its validation and reconciliation with the conflicting evidence of previous studies requires a comparison of volatility forecasts in order to find that the inconclusive results are due to discrepancies in predictive accuracies, which thus became a secondary hypothesis of this study.

work is plagued by data-mining (that is, if researchers set out to find deviations from rational pricing by running numerous regressions, ultimately they will be successful). However, much empirical work has confirmed the evidence out-of-sample, both in terms of time-periods as well as cross-sectionally across different countries. Finally, it is often claimed that behavioural finance presents no unified theory unlike expected utility maximisation using rational beliefs. This critique may well be true at this point, but traditional risk-based theories do not appear to be strongly supported by the data. Thus, it appears that there is a strong case to build upon some theories that are consistent with evidence, than theories based on rational economics whose empirical support appears quite limited. Indeed, a ‘normative’ theory based on rational utility maximizers cannot be construed as a superior alternative to behavioural approaches merely because it discusses how people should behave. If people do not behave in this way, this approach has limitations in helping us understanding financial phenomena.” (p. 13)

As a next step the tests have to be operationalized. Indeed, the model of Merton (1987) assumes a two-period setting with deterministic and known volatilities and homogeneous beliefs on the parameters of the model. In this way the model concentrated on predicting equilibrium prices (and therefore, in the two-period setting, returns) and abstracted the parameters that were easier to estimate like variances. The model also did not make any suggestions about what the investment horizon for the decision-makers is. Therefore, the operationalization of the theoretical model requires making decisions on how to apply the model to the empirical setting. The first choice is the horizon of the tests. The model of Merton (1987) is set up as a two-period economy where at time $t = 0$ investors make allocation decisions, and at the end of the period, time $t = T$, investors collect the uncertain dividend (positive or negative) on their investments. Thus, the predictions of the model should be valid at any frequency – daily, weekly, monthly, quarterly, yearly, or any other frequency. In practice daily, weekly and monthly frequencies tend to be the most common ones. Tests involving all time frequencies would increase significantly the scope of this research, especially given that different frequencies are associated with different amounts of data and face different challenges. For example, five years of data at a monthly frequency imply 60 realisations, which limits the types of models that can be estimated with reasonable accuracy. On the other hand, five years of daily data translate into over 1250 realisations, which allows estimation of much more complex models. High frequencies (daily and intra-daily), however, are exposed to market microstructure effects that may affect tests trying to match daily returns to risk. Market microstructure noise refers to short-term deviations between reported prices and the actual fundamental or equilibrium prices due to the way stock markets operate, e.g. bid-ask spreads, depth of the order book, rounding errors, screen fighting or other considerations. Whatever the cause, that noise obscures the ‘true’ market prices and returns; for example, a large buy bid may not be executed in one transaction but may take a couple of trades with increasing prices, which could result in reporting higher daily return, whereas in fact the prices subsequently revert to the equilibrium value. Therefore, the estimation methodologies suitable for monthly data may not be well suited for drawing conclusions with daily data. In our problem most of the controversial findings are based on tests using monthly data, and therefore we decide to concentrate on that frequency as well, and at that frequency the support for the prediction of the Merton model is weaker.

Another decision in the operationalization of the model concerns the selection of measures of idiosyncratic risk. The models of Levy (1978) and Merton (1987) associate idiosyncratic risk more or less directly with the distribution of the company-specific shocks,

and more specifically, with the standard deviation of that distribution. The model of Levy (1978) considers investors' individual portfolio optimisation problems in the mean-variance efficient optimisation framework where idiosyncratic risk is naturally associated with the standard deviation of idiosyncratic shocks in the linear factor model. In the model of Merton (1987) idiosyncratic risk is introduced as a random shock on company cash flows in an economy characterised with quadratic utility function, so that again the standard deviation of the shock distribution is the natural measure of idiosyncratic risk. Nevertheless, we should also recognise that those models are an abstraction, and individual risks may not be linear, while utility functions need not be well approximated by quadratic utility function. Concerning the first objection we shall note here that in practice exposures to risks are not known to investors accurately, but are learned from many factors. This process of learning and mixing market data with soft information and individual perceptions to reach an assessment of the riskiness of an investment may suggest that qualitative factors may be important in that respect, a consideration that we discuss in somewhat greater detail in the following subsection.

Concerning the second objection – that investor utility may not be approximated well by quadratic utility, we note that such considerations may give rise to significance of other idiosyncratic risk characteristics like skewness, as suggested by Barberis and Huang (2008), or tail risk, as reported by Huang, Liu, Rhee and Wu (2012). The options to pursue these directions were limited by the choice of the monthly frequency for the tests. We have nevertheless attempted to estimate tail risk from the parameters of the distribution of idiosyncratic innovations from GARCH models, as well as sample own skewness, co-skewness, and co-kurtosis, but we found that in view of the short series these analyses were not conclusive and are not reported here. Therefore, we concentrate on idiosyncratic volatility (standard deviation) as a principal idiosyncratic risk measure in this study, while recognising that other qualitative and quantitative risk characteristics could be added in the analysis to improve accuracy.

We will defer the discussion of the tests employed in this study to the subsequent subsection on statistical tests. At this juncture we can only note that the step involved selecting methods to compare predictive accuracy from different models, e.g. though loss-functions, Mincer and Zarnowitz (1969) regressions, or other tests. Similarly, the choice of test of the link between idiosyncratic risk and returns allowed the use of portfolio spreads, Fama and MacBeth (1973) regressions, Generalised Least Squares (GLS), or Generalised Method of Moments (GMM). We will motivate our choices in the subsection dedicated to that

topic.

Finally, in view of the evidence contradicting the theoretical model we were invited to either revisit our operationalization of the model, or reject the model altogether. We found that the introduction of the mean-reverting level of volatility was in fact a plausible rationalisation in the context of dynamic, stationary volatilities and it resulted in overall confirmation of the theoretical models.

Wrapping up the discussion, we find that the research problem can be approached by an array of inductive and deductive research strategies, each having its advantages and limitations. Our first reason to prefer deductive over inductive research is the material risk of data mining, which could result in a discovery of spurious correlations that do not generalise outside of our sample. The second driver for our preference for the deductive approach is the goal to reach conclusions that generalise outside of the studied sample, e.g. in other time periods or other marketplaces, institutional arrangements, market segments, as well as to anticipate the effects of market infrastructure changes that could result in lower transaction costs (technological advances), shorter planning horizons (e.g. algorithmic trading), or improved transparency (e.g. robo-advisors). We see these as attractive aspects of deductive research, and therefore we choose for our research a deductive approach, though we do adapt our operationalisation of the theoretical models if required by the available evidence.

Table 3: Deductive research workflow

Stages in the deduction process	Actions taken	Example: application to the present research
Context	Read and consider	Review of the factors affecting stock returns revealed contradictory evidence, e.g. Ang et al. (2006); Bali and Cakici (2008); Fu (2009), as well as the evidence of wide under-diversification of investor portfolios.
Theory	Select a theory or set of theories most appropriate to the subject under investigation.	Review of the theories purporting to predict the phenomenon, e.g. CAPM, the models of Levy (1978) and Merton (1987), and select between those and alternative explanations, e.g. behavioural models.
Hypothesis	Produce a hypothesis (a testable proposition about the relationship between two or more concepts).	We hypothesise that there is sufficient under-diversification in the market to produce positive correlation between idiosyncratic risk and stock returns. We hypothesise that the conflicting evidence in previous research is due to discrepancies in the operationalization of the tests.
Operationalize	Specify what the researcher must do to measure a concept.	The theoretical model is an abstraction (two period, homogeneous beliefs) and can be employed in various horizons (daily, monthly, etc), different measures of idiosyncratic risk (e.g. volatilities, tail risk, dispersion of analysts' forecasts, etc).
Testing by corroboration or attempted falsification	Compare observable data with the theory. If corroborated, the theory is assumed to have been established.	Test the differences in forecast accuracy through Mincer and Zarnowitz (1969) regressions. Test the correlation between volatilities and returns using different tests (portfolios vs individual securities, portfolio spreads vs Fama-Macbeth vs Generalized Method of Moments).
Examine outcomes	Accept or reject the hypothesis from the outcomes.	Analyse the results from the tests above.
Modify theory (if necessary)	Modify theory if the hypothesis is rejected.	Adapt the operationalization of the theory, e.g. exclude idiosyncratic tail risk as insignificant and introduce the mean-reverting level.

Source: adapted by the author from Table 2.1, p. 17 in Gray (2014)

3.2.3. Quantitative and qualitative research

Research approaches can be broadly classified into quantitative and qualitative, depending on the type of interpreted information. Examples of problems of qualitative nature are the analysis of meanings expressed in words, images or sounds, analysis of non-standardised data requiring classification into categories, or analyses conducted through the use of conceptualisation; see Table 13.1, p. 482 in Saunders et al. (2009). Quantitative research, on the other hand, applies to cases where meaning is extracted from numerical, standardised data and analysis is conducted through the use of diagrams and statistical methods.

It is difficult to give a precise definition of what qualitative research means because the scope of research approaches is quite wide. Instead, Yin (2011) identifies five key traits of qualitative research:

- “1. Studying the meaning of people’s lives, under real-world conditions;
2. Representing the views and perspectives of the people [...] in a study;
3. Covering the contextual conditions within which people live;
4. Contributing insights into existing or emerging concepts that may help to *explain* human social behavior; and
5. Striving to use *multiple sources of evidence* rather than relying on a single source alone.” (p. 7-8)

Qualitative research methods are often beneficially used in studies that interpret social phenomena in order to reveal their causes and effects. Such interpretative studies recognise that unlike the natural world, the social world is constructed by humans and has valid subjective description. Qualitative studies often use inductive⁵⁷ reasoning in order to uncover the causes and effects of the studied phenomena using tools like case studies, interviews, focus group discussions, in-depth interviews, observations and field trips. In contrast, quantitative studies interpret numerical, standardised data; quantitative methods are preferred for confirmation of positive theories and deductive arguments that extract meaning from

⁵⁷ The use of the deductive approach in qualitative research is subject to some debate. Some authors support its application, but there are also arguments against that. For example, Bryman (1988:81) (quoted in Saunders et al. (2009), section 13.4) argues as follows: “The prior specification of a theory tends to be disfavoured because of the possibility of introducing a premature closure on the issues to be investigated, as well as the possibility of the theoretical constructs departing excessively from the views of participants in a social setting.”

numbers through measurement, statistical analyses and test of hypotheses.

An empirical analysis of the link between idiosyncratic risk and returns could be tackled with both qualitative and quantitative methods. For example, in-depth interviews with decision-makers could be conducted in order to understand how they form their portfolios in terms of balancing the various criteria – some of which qualitative – and making allocation decisions. Similarly, how decision-makers perceive of risks could also be fruitfully examined through such qualitative methods. In business research one often deals with notions like corporate image or social responsibility, which may not have a standardised quantitative measurement. Even if those concepts could be quantified, it is often impossible to collect such data in order to perform a quantitative confirmation of the developed theory, which could limit the generality of the developed model and its confirmability.

Another difficulty in implementing qualitative research in the field of asset pricing is the large number of biases found in investors. This problem is not unique to finance, but nevertheless seems more acute in the field of asset pricing, possibly because the financial impact of decisions is easier to quantify *ex post* and asset managers need to make routinely predictive decisions in a random environment, which incites the activation of psychological defence mechanisms to handle disappointments. Therefore, while overconfidence may be a common trait in humans, it is likely to be more pronounced in the financial decision-making process. For example, Aronson (2007), Chapter 2, considers eleven psychological biases and mechanisms that give investors illusory confidence over the validity of their statements: (i) overconfidence, optimism bias; (ii) self-attribution bias; (iii) illusion of control; (iv) knowledge illusion; (v) biased second-hand knowledge; (vi) representativeness heuristic bias; (vii) sample size neglect; (viii) illusory trends and patterns; (ix) hindsight bias; (x) confirmation bias; (xi) illusory correlation. Other classifications also exist, and their findings are similar. Such behavioural biases are also recognised by industry practitioners and associations. For example, Pompian (2016) recognises a total of 20 behavioural biases affecting investors, grouped into behavioural persistence, information processing, and emotional biases. The presence of significant behavioural biases in investors would make very difficult the separation of actual behaviours from biased rationalisations, and would be a formidable obstacle in generalising text-based qualitative information into a test of an asset pricing relation.

In view of the interpretative difficulties of verbal information and limitations on result generalisation, we opt not to pursue a grounded theory study where methods like depth interviews are employed to collect data on how investors assess stock risk and whether and

how they incorporate idiosyncratic risk into their decision-making process, and the collected information is then generalised into a theory of the link between idiosyncratic risk and investment allocation. In the data collection step, the interviewees may tend to demonstrate overconfidence and may produce misleading results. In the theory-development step, we should recognise that the stock market needs to clear so that the prevailing prices should equate supply and demand, so that our theory would need to take into account how aggregate demand for all assets is affected by the developed model.⁵⁸ Furthermore, a small number of interviewees would limit the ability to generalise the results to the whole market and sample period, and could undermine the goal of this research to uncover the causes for the mixed evidence on the correlation between idiosyncratic risk and returns.

Returns and portfolio allocations are essentially quantified in terms of continuous measures, which suggests that quantitative research might be well-suited in explaining the phenomena at hand. Moreover, the field of asset pricing has been extensively researched in the past sixty years, which, combined with the extensive availability of data on stock returns, could allow conducting tests of existing theories on the entire population, rather than just on a small sample. This is an important consideration because those theories predict a stochastic causal relation between risk taken and return earned. The probabilistic nature of the causality suggests that there is substantial probability that a flawless decision-making process could nevertheless result in loss-making investments purely by chance, and therefore general conclusions about the link between idiosyncratic risk and returns require confirmation in a large sample.

Overall, in this study we choose to use quantitative methods. Our first argument for that is that the predicted phenomenon (market returns) is inherently quantitative in nature and can be evaluated with high precision. Secondly, we are concerned about the reliability of many qualitative methods in view the evidence of a number of psychological biases of investors, documented in existing research. A decision to choose quantitative over qualitative research approach, however, comes at the cost of loss of detail. Considerations about the specific decision-making process and multiplicity of investor objectives that could be uncovered or examined through qualitative research are sacrificed. We recognise those

⁵⁸ If investors chose to reduce their investment in, say, stock A, they still need to invest the remaining amount into other stocks, say stock B, and the risk-free asset. This illustrates a core proposition of general equilibrium models where changes in the demand for one asset (stock A) impacts all the prices of all other assets (stock B and the risk-free rate, in this simple example).

limitations of our research design, but having in mind our objective to revisit the empirical support of existing theories and reconcile the conflicting evidence in previous research, we have opt for the use of quantitative research design.

3.2.4. Statistical methods

In the last subsection we reasoned that quantitative methods are better suited to the intended deductive approach in a manner that could allow reaching general conclusions about the validity of the theoretical model. There is an array of methods that can be used in quantitative research, and in this section we shall motivate the use of the quantitative methods employed in our study.

There are various classifications of quantitative methods, e.g. descriptive, exploratory, experimental, or statistical. Descriptive research aims to identify the characteristics of the explored phenomenon. For example, it could employ surveys and questionnaires to gather information about how investors assess the risk of their holdings and make allocation decisions. A study of this type could mix both quantitative and qualitative information in order to answer such a question. Information on risk assessment methodologies and the incorporation of idiosyncratic risk into portfolio allocation decisions could be collected through questionnaires. That method also has the advantage that the questionnaire could be structured to collect both qualitative information (e.g. through open questions) and quantitative data. However, the method is exposed to problems similar to those of depth interviews and qualitative research: the presence of behavioural biases would suggest the use of long questionnaires that could detect inconsistencies and behavioural biases, which would limit the response rate and representativeness of results. Furthermore, in view of the evidence of changing investment patterns presented by McLean and Pontiff (2016) it would be difficult to generalise the results of that study to explain patterns in the past. Finally, it would be difficult to reconcile the conflicting evidence from previous studies in that approach, inasmuch as their differences are only indirectly related to information that could be gathered in such a manner, e.g. number of assets in portfolios and attitudes towards portfolio diversification. The dependent characteristic that needs explanation is return, and descriptive methods would not be able to estimate reliably what premia investors require for assuming different exposures to idiosyncratic risk.

Exploratory research could also be employed in a related context to classify investments by risk and return. For example, we could collect or estimate various

characteristics associated with idiosyncratic risk, e.g. frequency and quality of disclosures, variability of cash flows or stock prices, co-movement with market, sector, indebtedness, company age (life cycle), or any other characteristic. We can then seek to classify companies in various categories, seeking to explore how each category differs from the others in terms of characteristics and returns. For example, we could explore whether NYSE-listed companies on average earn more than Nasdaq-listed companies. Various methods could be employed in such a study, e.g. clustering (for unsupervised grouping of companies into clusters for the purposes of exploring how those clusters differ in terms of average returns), neural networks (for predicting growing companies from a set of characteristics), analysis of variance (ANOVA, for comparison of returns for various categorical dimensions, e.g. sector or market). Such an approach could be useful to develop new hypotheses for the factors affecting returns. On the other hand, such an approach could be questioned in terms of its external validity on the same grounds as data mining that we discussed in the inductive research section.

Experimental designs offer another possible approach for addressing the research question. In the present context a number of experimental settings could be envisaged. For example, one can devise various lotteries and offer them to participants in the experiment in order to confirm whether choices are consistent with those of the expected utility theory and identify behavioural patterns, e.g. preference for positive skewness. Alternatively, one may set up a virtual trading platform and explore how participants form their portfolios in a controlled setting. Yet another possibility could be to perform simulations that re-create the theoretical setting of the different models and explore how premia change with a number of securities in individual portfolios, with the relaxation of some of the assumptions, with inclusion of parameter uncertainty (e.g. via Bayesian learning), etc. For example, Allais (1953) demonstrates in experimental setting a violation of the independence axiom.⁵⁹ Smith

⁵⁹ The Allais (1953) paradox is a well-documented violation of the independence axiom. In his experiment individuals are asked to select between two pairs of lotteries. The first lottery v_1 promises a certain payoff of 100 million francs, whereas the alternative lottery v_2 offers payoffs of 500mn, 100mn and 0 with respective probabilities 0.1, 0.89, and 0.01. He finds that most people prefer v_1 over v_2 , $v_1 \succ v_2$. The other pair of lotteries are u_1 , paying 100 mn and 0 with probabilities 0.11 and 0.89, and u_2 , paying 500mn and 0 with probabilities 0.10 and 0.90. He finds that most people prefer u_2 over u_1 , $u_1 \succ u_2$. The paradox was that there is no utility function $U(\cdot)$ that would satisfy these choices. The first pair implies that $U(100) > 0.10 U(500) + 0.89 U(100) + 0.01 U(0)$, so rearranging the terms we obtain that $0.11 U(100) > 0.10 U(500) + 0.01 U(0)$. On the other hand, the second pair of lotteries suggested that $0.11 U(100) + 0.89 U(0) < 0.10 U(500) + 0.90 U(0)$, and thus after a rearrangement yields $0.11 U(100) < 0.10 U(500) + 0.01 U(0)$, contradicting the inequality from the first pair. This shows that the independence axiom may be violated in

et al. (1988) observe trades in a series of simulated asset markets and document the emergence of bubbles in most of them, although the probability of bubbles decreases with trader experience. Asparouhova et al. (2016) simulate the equilibrium in an extension of the Lucas tree economy⁶⁰ and find that trading substantially improves intertemporal consumption smoothing, but prices remain volatile, similarly to real markets. Such studies could provide valuable information on the topic, but also come with a set of challenges. Experimental designs that elucidate deviations of the axioms of the expected utility theory are questioned as concentrating on deviations that emerge in contrived settings or small lotteries and do not necessarily translate into similar behaviour in larger, real-life behaviours. At any rate deviations from the assumptions of decision-making under uncertainty would have implications far beyond the problem at hand, and such deviations would need to be confirmed in a range of other settings, that go well outside the scope of this study.

Similarly, simulations from the model economy would ultimately be determined by the quality of the underlying model and its operationalisation, and assessment of their validity would necessitate development of new approaches for their validation, e.g. in terms of response to changes of volatility in a simulated economy and in actual markets. Such exploration could be a useful contribution to experimental designs, but again goes beyond the goals of this thesis and its scope.

In the same vein, the effect of parameter uncertainty could be significantly contributing to the actual outcomes, and in fact we find evidence in favour of the hypothesis. We find that whereas one-month idiosyncratic volatility forecasts are not useful predictors of returns, the mean-reverting ones are. This could be consistent with investor uncertainty over future volatility, augmented with longer holding periods or trading costs that make portfolio rebalancing costly. This highlights the difficulties in employing simulation designs and generalising their results for the whole market: the outcomes would depend on many

practice, and the choice between lotteries could depend on their context; in particular, the first pair offers the choice between a certain win and a speculative bet, and most people tend to go for the certain sum. The second pair offers the choice between two speculative lotteries and most people tend to go for the one with the larger payoff in the unlikely event of winning. Following Allais (1953), many more tests were conducted and these confirmed that in various combinations of lotterys people tend to make choices that violate the independence axiom, like certainty and isolation effects (Kahneman and Tversky, 1979) and the context (framing) effect (Hershey and Schoemaker, 1980).

⁶⁰ Lucas (1978) proposes a model of exchange economy populated by an infinite number of identical individuals, each of which is endowed with a never-perishing tree that yields a random crop of apples (dividends) each year. In that setting he demonstrates that the price of the trees equals the present value of future crops discounted with a stochastic discount factor.

assumptions that are incorporated in the simulation – e.g. holding period, transaction costs, parameter learning – and the external validity of the results is therefore difficult to ascertain.

In this study we employ statistical methods in order to address the research questions at hand. Such methods can be used with any type of numerical data, but are particularly well suited for studies that investigate continuous data in terms of correlation and stochastic (probabilistic) causation. Returns are in practice a continuous random variable, therefore such methods allow the estimation of expected mean returns or return quantiles conditional on a set of categorical, ordinal, or numerical variables. This could allow the estimation of spreads in returns between different clusters of companies in a manner that recognises the correlation of the explanatory factors and estimates the contribution of each independent variable on the conditional return moments and quantiles. For example, idiosyncratic risk could be associated with higher beta, lower liquidity, younger company age, smaller size, or smaller book value of equity. Some of the simple descriptive or exploratory methods do not allow control for such a correlation and consequently it is difficult to attribute observed differences in returns to each of the correlated factors. Other methods (e.g. neural networks or some non-parametric methods) may be better suited to handle non-linear dependencies,⁶¹ but the cost is obscuring the contribution of each independent variable on stock returns, making the model behave like a “black box”. Instead, we choose to conduct our analyses using linear regression models, where dependent variables are hypothesised to be linear functions of a set of explanatory variables. That approach offers transparent interpretation of regression parameters and directions of co-movement, and still could accommodate complicated models, including polynomial ones.

Therefore, in our study we chose to use regression models to infer the relationship between returns and risk characteristics like beta, size, liquidity, and idiosyncratic risk. The choice is closely related with the choice of deductive approach, and in particular our concerns about external validity, generalizability and truth-preservation. In this way we can expect that our conclusions will be stable in time, markets, institutional settings, market infrastructures, and our conclusions would not be overturned or be significantly amended from new evidence. Similarly, our objective to prevent data mining and model overfitting compelled us to use regression analysis instead of non-linear or non-parametric models.

⁶¹ The correlation is essentially a measure of linear dependence. It is possible for two variables to be dependent in a non-linear fashion, and yet their correlation could be zero.

3.2.5. Regression models

3.2.5.1. Correlation and causation

The use of statistical models requires a set of explanatory variables (categorical, ordinal, or numerical) that are hypothesised to be affecting the conditional distribution of the dependent variable. The concerns about external validity and how well the results generalise out of our sample require the use of as large a sample as possible. This consideration forced us to leave out characteristics, which could predict the risk profile of the company, but were not publicly available or were available for only a fraction of all companies or subsets of years. Therefore, we had to give up on use of some measures of idiosyncratic risk and make do with what could be readily obtained for a sufficient number of companies. For example, the variability of net income or sales could be a better proxy of idiosyncratic risk that is not affected by “noise trading”, but due to the absence of such information for a representative sample of companies (especially for the smaller ones), we could not implement such an approach. Similar considerations do not allow the use of option-implied volatilities, because these are available only for recent years and for parts of the market (e.g. larger companies) that may not be representative for the entire cross-section. Similarly, availability of information on individual transactions could provide better measure of breadth of the investor base, and hence better measure of undiversified idiosyncratic risk, but such information was not available to us. Information on other characteristics that could affect the conditional mean returns like investor profile (e.g., the share of institutional ownership) could be useful in clarifying whether those investors were likely holding undiversified portfolios and would be seeking a risk premium for idiosyncratic risk, but again such information was not available.

Therefore, the operationalisation of the theoretical model for statistical testing requires the estimation of risk characteristics using mostly publicly available data: stock prices, capitalisation, book value of equity, traded volume. Other characteristics had to be estimated from this limited set of data, which is clearly one of the limitations of the approach. Idiosyncratic risk is estimated from the limited available information on the total stock returns (daily and monthly) and the available information of market excess returns, the Fama–French factors, and, in some cases, market momentum. Thus, idiosyncratic risk is reduced to just the changes of prices that were not attributable to some market trend (market return, Fama–French factors, or momentum). This definition of idiosyncratic risk is much narrower than what could be considered in other designs, but has the advantage that it can be estimated for almost all securities and thus enhances the external validity (generalisability) of our results.

Statistical analyses can be performed in terms of correlation, but our context requires more – establishing the causation from risk to returns. For example, we can note that if some characteristic (e.g., size) is correlated with higher average return, this does not imply that necessarily size is a characteristic that is valued by investors. Instead, it might be the case that smaller companies have less mature business models and diversified cash flows, which makes them more vulnerable to economic shocks. In that case, the correlation between size and returns would not imply that size is ‘causing’ higher returns, only that they are correlated. In general, causality is a concept that is difficult to implement in statistical studies. Probably the most popular statistical test of causation is the Granger (1969) causality test. In his approach, one time series (X_t) Granger-causes another one (Y_t), if past history of the first series ($X_i, i = 0 \dots (t - 1)$) together with the past series of the predicted series itself ($Y_i, i = 0 \dots (t - 1)$) predicts Y_t . Nevertheless, it does not follow that X_t truly causes Y_t .⁶² Therefore, our design has certain limitations in establishing causality *per se*, without the context provided by the theoretical models, from which the relationship is developed – Levy (1978) and Merton (1987). To address these concerns, as well as look-ahead bias, we use not the realised returns in a given month, but the expected values based on the past history. Nevertheless, it does not follow that those statistical analyses necessarily establish causality between idiosyncratic risk and volatilities. Instead, they establish correlation in a setting that aims to reduce look-ahead bias. The causation is only inferred from the fact that this correlation pattern is predicted by the theoretical model; it is still possible, however, that idiosyncratic volatilities serve as a proxy for company exposure to some other risk factor, and the underlying model is overall not valid. We aim to mitigate those concerns by implementing a host of robustness checks, but the risk cannot be eliminated altogether.

3.2.5.2. Factor model estimation

The choice of specific statistical methods depends on a number of factors, like the types of variables available, the types of hypotheses tested, assumptions and powers of the various tests. Most of the explanatory variables available to us in this study (either retrieved from external data sources, or calculated by us) were continuous, which allowed significant flexibility in selecting statistical methods. In this phase we have to make three principal choices. The first one is the selection of methods to estimate idiosyncratic returns (also interchangeably referred to as shocks or innovations) and methods to forecast the expected

⁶² “Post hoc ergo propter hoc” fallacy (Lat. “after this, therefore because of this”).

idiosyncratic risk. The second one is selecting methods to compare the predictive accuracy of the alternative forecasts. The last one is a method to measure the impact of idiosyncratic volatilities on realised excess returns. In this subsection we shall explain the principal options available to us, while the details of their implementation are provided in the remaining sections of this chapter.

The first part of the decision concerned what sort of model could be used to forecast returns. Idiosyncratic risk is an abstract concept that is difficult to define. In the model of Merton (1987) it is simply a random deviation of end-of-period production of each firm that is uncorrelated with the rest of the investees. In our setting this concept requires an operationalisation, where idiosyncratic risk is the difference between the observed actual return on the stocks and the expected return on the stock, estimated from some model, the parameters of which are estimated from past periods. It is possible to select some non-linear model for stock selection, e.g. Levin (1995), and use its forecasts to separate actual returns into expected and unexpected components, we acknowledged that such an approach is rarely pursued in literature and would obscure comparability of our results with other research in the field. At any rate it should be emphasized that linear asset pricing models can accommodate many non-linear situations⁶³ and that assumption should not be considered a substantial limitation.

Ross (1976) proposes an alternative to the CAPM reviewed previously – the Arbitrage Pricing Theory (APT). He assumes that N factors drive returns in linear fashion, i.e.

$$r_{i,t} = \alpha + \sum_{n=1}^N \beta_{n,i} f_{n,t} + \varepsilon_{i,t},$$

where $\varepsilon_{i,t}$ are mutually uncorrelated and have expected value zero and finite variance, i.e. $\mathbb{E}\varepsilon_{i,t} = 0$, $\mathbb{E} < \infty$, $\mathbb{E}\varepsilon_{i,t}\varepsilon_{j,t} = 0$. Furthermore, it is assumed that the expected values of factors equals zero ($\mathbb{E}f_{n,t} = 0$), so that prices are affected only by the unexpected factor realisations. Notably, the model does not assume that the factors are mutually independent ($\mathbb{E}f_{n,t}f_{m,t}$ need not be zero), and even does not require factors to be independent with the errors ($\mathbb{E}f_{n,t}\varepsilon_{i,t}$ need not be zero) or have finite variances. Ross argues that in an efficient market one should not be able to construct an arbitrage zero-investment portfolio.⁶⁴ In

⁶³ E.g. it is allowed that one factor is a square of another one or product of two others.

⁶⁴ To prove the existence of an arbitrage-free equilibrium he assumes the following: 1) There is at least one asset with bounded losses; 2) There exists at least one investor who is uniformly less risk-averse than some constant relative risk-averse agents, and who believes that the assets are generated by the linear model stated above, and who is not asymptotically negligible; 3) All agents are risk-averse and hold the same expectations; 4) the aggregate

empirical applications the factor model takes the form

$$E_i - \rho \approx \beta_{1,i}\gamma_1 + \beta_{2,i}\gamma_2 + \dots + \beta_{N,i}\gamma_N,$$

where γ_j is the risk premium on a portfolio that has unit exposure only to factor j ; if E_j is the expected return on a portfolio with unit loading only to factor j and zero loadings on all other factors $f_{i,i \neq j}$, then $\gamma_j = E_j - \rho$.

The APT does not stipulate a list of factors; it is therefore useful to identify the criteria that should be met by candidate factors. The structure of the model presents immediately two eligibility criteria: firstly, the assumption of zero expected factor value implies that factors affect prices through their unexpected realisations. Secondly, the factor should be pervasive in the sense of affecting sufficiently many securities through the linear relationship above, or else randomly constructed portfolios would tend to diversify it away, producing portfolios with no exposure to that factor. Many macroeconomic factors appear suitable candidates, and Chen et al. (1986) investigate how various macroeconomic factors could explain asset returns. They conclude that the following factors significantly affect stock returns: industrial production (led by one period to make it contemporaneous with asset prices); changes in risk premium (the spread between the yield on portfolio of ‘Baa’ and lower grade bonds, and long-term government bonds); twists of the yield curve (the spread between a short and a long point of government yield curve); unanticipated inflation (the difference between actual inflation and expected inflation estimated by macro-econometric model); changes in expected inflation; the last two inflation-related factors were reported to be more significant in more turbulent periods characterised with volatile inflation.

Macroeconomic models are useful for modelling macroeconomic influences on portfolio values, and various hybrid models (combining macroeconomic series with other data) are implemented in the financial industry. For example, Northfield (2013) combines the macroeconomic factors of Chen et al. (1986) with other series like oil prices, housing starts, exchange rates, and five statistical factors in order to model portfolio macroeconomic risks.

Macroeconomic factor models are occasionally criticised for poor explanatory power⁶⁵. Furthermore, the classic model of Chen et al. (1986) highlights the implementation difficulties. Thus, while stock prices are available at intra-day frequencies, many economic series (e.g. industrial production, inflation, gross national product) are available only at

demand for all assets is non-negative; 5) expectations are uniformly bounded. Under those assumptions he proves that there exist ρ and β_i such that $\sum_{i=1}^{\infty} (E_i - \rho - \beta_i\gamma) < \infty$, with ρ equal to the risk-free rate, if a risk-free asset exists in the market.

⁶⁵ See, for example, Connor (1995)

monthly or even quarterly frequencies. The flow series may entail a lag (hence industrial production needs to lead one period or more); moreover, information when the values of the macroeconomic indicators are released to the public and incorporated in asset prices as well as any subsequent revisions of the indicator are usually hard to obtain. What matters for APT are unanticipated factor realisations, hence one needs to implement a model of the rationally expected values (like the expected inflation, in order to derive surprise inflation), which inherently entails model risk. Furthermore, part of the information of the economic environment would be acquired by investors either by direct observation, or through other proxy variables, so it is unclear what part of the effect of the economic datum should be realised during the period of the slowdown and what part should be realised upon the release of the economic data. Finally, the impact of macroeconomic releases need not be flat, and large surprises may trigger more than proportional response relative to small surprises. Because of these limitations, arbitrage models with trading factors are often preferred over macroeconomic models.

Therefore, in our study we employ a linear factor model with traded factors in order to separate the observed returns into systematic and idiosyncratic components. In different contexts we use a varying number of factors, depending on the context. The exact list of factors used is discussed in the next section.

Estimation of the parameters of the linear factor models is deceptively straightforward. The common practice in the literature is to use the Ordinary Least Squares ('OLS') estimator that minimises the sum of the squared errors. The OLS estimator has the desired property of being the best linear unbiased estimator of model parameters, however, since the objective norm is squared error, differences between the observed and fitted values are squared and thus outliers could significantly affect the estimates. Another possible approach is to employ a robust estimator of regression coefficients. Such estimators are, for example, the Least Absolute Deviation ('LAD')⁶⁶, the M-estimator⁶⁷, or the Trimmed Regression Quantile⁶⁸, to

⁶⁶ The LAD estimator minimises the sum of the absolute values of the errors, i.e. solves $\min_{\beta_i} \sum_t |\varepsilon_{it}|$

⁶⁷ The M-estimators minimise the sum of errors scaled by some factor σ and weighted through some function ρ , i.e. they solve $\min_{\beta_i} \sum_t \rho(\varepsilon_{it}/\sigma)$. There are many choices for ρ , for example the function for the Huber's M-estimator is

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| < c \\ c|x| - \frac{1}{2}c^2 & \text{if } |x| \geq c \end{cases}$$

name just a few of the options.⁶⁹ In practice, however, non-OLS estimators for factor models of asset returns are exceptions rather than the rule, possibly because the asymptotic theory for these estimators is less developed compared to estimators like OLS and Maximum Likelihood, which hinders hypothesis testing. Concerning unbiasedness of the LAD estimator, Gray et al. (2013) observe that LAD beta estimates are systematically below the corresponding OLS estimates. They estimate the CAPM regression for all securities in the Australian market from January 1976 to May 2012 using both the OLS and the LAD estimator. They observe that if an estimator of beta is unbiased, then the value-weighted average of individual betas with respect to a specially constructed index with constant weights should equal 1 when average is taken for all constituents of the index. They find that while that was indeed the case for the OLS estimator, the LAD averages below 1, thus suggesting possible downward bias of the LAD estimator.

Having regard for concerns that robust estimators might be biased and in order to facilitate comparisons with the existing literature, we opted for using the OLS estimator to calculate the parameters of the linear factor models. Nevertheless, we believe that the topic of the estimation method could be a worthwhile direction for future research. We note that in some cases the betas estimated through OLS are well outside the corridor 0 to 3, which in itself is already quite wide. This indicates that in some cases there may exist larger idiosyncratic shocks in one single month, which then result in the factor model attempting to reconcile the factor loadings with the extreme return. In the case of simple OLS idiosyncratic volatility⁷⁰, that could result in OLS volatility being lower than the volatility when a robust model was being used.

Wrapping up the discussion in this subsection, we choose to use traded factor models for asset returns in order to split returns into systematic and idiosyncratic components. The models shall be fitted using Ordinary Least Squares (OLS) estimator.

3.2.5.3. Selection of volatility models

The estimation of volatilities was a principal methodological decision in our research.

⁶⁸ The TRQ estimator is a weighted average of different beta quantiles. In general, $\beta_\alpha = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \beta(\theta) d\theta$. In practice this works by fitting some finite number of quantile regressions and then calculating a weighted average of the estimates.

⁶⁹ See Chan and Lakonishok (1992)

⁷⁰ We define the idiosyncratic volatility measures later in the chapter. In this case the OLS volatility is simply the standard deviation of the error term from the fitted factor model.

We hypothesise that differences in precision across alternative forecasts of volatilities result in the conflicting findings in the literature. Therefore, we aim to use forecasts that meet three criteria. Firstly, we want our forecasts to be consistent with existing research, so that our study could reconcile the conflicting evidence, rather than contribute to the confusion by offering yet another estimator. Secondly, we aim our estimators to be transparent inasmuch as possible, so that the obtained results could be intuitively interpreted. We want to avoid a situation where the models operate as black boxes and the results can be interpreted principally in terms of statistical significance. Finally, we want our estimators to be sufficiently representative for the available forecasting methods, rather than compare variants of the same or very similar methods.

The simplest estimator of idiosyncratic risk variance is the OLS estimator, which is simply the standard deviation of the residuals from some factor model:

$$\sigma_{ols,t}^2 = \sum_{k=1}^T \frac{(\varepsilon_{i,t-k})^2}{T-df},$$

where $\varepsilon_{i,t-k}$ denotes the idiosyncratic residual from the factor model for stock i in month $(t-k)$ obtained from a regression estimated over the period $(t-T)$ until $(t-1)$. T is the actual number of months for which information is available; thus T is between 60 and 30 (the minimum number of months required for estimation); df was the residual degrees of freedom for the estimated model and equals the number of estimated parameters. The underlying assumption is that OLS variances do not change much from period to period as the estimates of month t volatility differs from month $(t+1)$ by just one return as one return exits the rolling window and being replaced by a new one as the window rolls forward. Thus, the realisation at period t does not affect the estimated expected volatility, i.e. the expected value is that as at the end of period $t-1$, or equivalently, at the start of period t . The benefit of that approach is the simplicity of the estimator and the filtering of random idiosyncratic innovations. On the other hand, in cases where idiosyncratic volatility permanently changed to a different value, that change would filter quite slowly in the volatility forecast.

Variances may change significantly from month to month. In that respect, it is useful to note another interpretation of the OLS variances and their relation to idiosyncratic returns. Observe that squared idiosyncratic return is in fact an estimator of the unobserved idiosyncratic variance.⁷¹ From that perspective, if idiosyncratic volatility consists of some

⁷¹ By construction (as residuals from OLS regression), the idiosyncratic residuals have mean

long-term, constant volatility and random deviations from the long-term volatility, then the moving average of the squared monthly idiosyncratic returns could be interpreted as a filter that removes the random deviations and extracts the long-term level.

However, the moving average filter also has a drawback – if idiosyncratic volatility is permanently increased at some point (a step increase), it would take five years (60 months) for that increase to filter entirely through the moving average filter. Therefore, a more sophisticated method for estimating idiosyncratic volatility could yield better results.

The OLS estimator was used in the early studies (in the 1960s and 1970s). Since then it has fallen out of favour as too unresponsive to changes. Nevertheless, we shall use it as one of our estimators for two reasons. Firstly, it serves as a benchmark that could provide perspective on how other models perform and give some intuitive feeling of how rigid the forecasts produced by alternative methodologies are. Secondly, in the foregoing paragraphs we interpreted the estimator as a filter, and in that sense it can be viewed as similar to other methods that can operate as filter as well. For example, the “exponentially weighted moving average”, which is a type of Integrated GARCH(1,1) and was used in the RiskMetrics methodology of J.P.Morgan/Reuters (1996), can be seen as an exponential filter, and thus an incremental improvement over the OLS filter. In our case it is useful to have the OLS benchmark, because it gives empirical perspective on the performance of GARCH(1,1) model.

The works of Engle (1982) and Bollerslev (1986) brought significant improvement of flexibility in terms of modelling of volatility evolution. Engle (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model, and his formulation was subsequently generalised by Bollerslev (1986), who formulated the Generalised Autoregressive Conditional Heteroscedasticity (GARCH(p , q) model). The simplest, but also remarkably successful version, is the GARCH(1,1) model:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (4)$$

The model always yields positive predicted variance when the three parameters are positive (ω) or non-negative (α , β), i.e. $\omega > 0$, $\alpha, \beta \geq 0$. When $\alpha + \beta < 1$, the model is covariance stationary⁷² and its unconditional⁷³ volatility is $\sigma^2 = \omega / (1 - \alpha - \beta)$. When $\alpha + \beta = 1$, the

zero. Then $Var(\varepsilon) = \mathbb{E}(\varepsilon^2) - (\mathbb{E}(\varepsilon))^2 = \mathbb{E}(\varepsilon^2)$.

⁷² A real-valued stochastic process $\{X_t\}$ is covariance-stationary if its expectation $\mathbb{E}X_t$ does not depend on t , and its k -th order autocovariance $\mathbb{E}[(X_t - \bar{X})(X_{t+k} - \bar{X})]$ is finite and depends only on k , but not on t .

⁷³ The unconditional volatility is simply the variance of the process, $var(X_t)$. The

model reduces to the Integrated GARCH(1,1) model of Nelson (1990). The motivation for that model is the observation that for financial time series, the sum of the two parameters ($\alpha + \beta$) is often quite close to 1. The unconditional variance of IGARCH is infinite, and implies a random walk of variances (i.e. every shock on volatility has permanent impact and never decays). IGARCH has the benefit of being a parsimonious specification and one that ensured that the estimated parameters are close to reality – certainly an asset if data is scarce. For this reason it is also the model that underlies the volatility modelling (exponentially-weighted moving average) of the successful RiskMetrics risk quantification specification of J.P.Morgan/Reuters (1996). However, the evidence of Fu (2009) of stationarity of variances for about 90% of all series convinces us to abstain from employing IGARCH as the principal specification of our study.⁷⁴

Since the introduction of the original GARCH(p, q), a number of alternative specifications were proposed, although GARCH(1,1) specification proved to be difficult to outperform out of the sample for one-step-ahead forecasts, as documented by Hansen and Lunde (2005). However, the model that actually gained traction in idiosyncratic risk tests is the Exponential GARCH model introduced by Nelson (1991). The model allows an asymmetric response to positive and negative innovations. Such asymmetric reactions are common in financial time series, hence the preference for that model in recent literature on idiosyncratic risk; it is also the forecasting model used in the studies of Fu (2009) and Malkiel and Xu (2004). A couple of slightly different specifications for the model exist. Using one of the specifications and changing the mean equation with the FFC gives:

$$(r_{i,t} - r_{f,t}) = \beta_0 + \beta_1(r_{M,t} - r_{f,t}) + \beta_2 r_{SMB,t} + \beta_3 r_{HML,t} + \beta_4 r_{MOM,t} + \varepsilon_{i,t}, \quad (5)$$

$$\varepsilon_{i,t} = \sqrt{\sigma_{i,t}^2} v_{i,t}, \quad (6)$$

$$\log(\sigma_{i,t}^2) = \omega_i + \sum_{k=1}^p \delta_k \log(\sigma_{i,t-k}^2) + \sum_{m=1}^q \alpha_m \{ \theta v_{i,t-m} + [|v_{i,t-m}| - E|v_{i,t-m}|] \}, \quad (7)$$

where ω_i , δ_k , α_m , θ , γ are the parameters of the EGARCH(p, q) model, $v_{i,t-m}$ is an i.i.d. random variable with mean zero and unit variance.

Fu preferred to employ the Exponential GARCH (EGARCH) model for modelling volatilities. The advantages of EGARCH over GARCH were that it ensured positive

unconditional variance at time t does not depend in the variance at time $t - 1$.

⁷⁴ We also calculated the expected forecasts from IGARCH but we found that GARCH(1,1) produced slightly more accurate results, which was an additional consideration in favour of GARCH(1,1) even through scarcity of data was an argument in favour of IGARCH.

variance⁷⁵ and allowed asymmetric response to shocks.

One downside of EGARCH models is that they have lighter tails compared to regular GARCH. If the GARCH(1,1) process is stationary, Mikosch and Starica (2000) showed that for distributions in the Frechet domain of attraction (Paretian tails),⁷⁶ the extremes of the GARCH process are heavy-tailed with extremal index $\theta \in (0,1)$, with expected length of clusters of high volatility in the normalised sequence equal to $1/\theta$.⁷⁷ Thus, Mikosch and Starica demonstrate that GARCH models allow for volatility clusters to exist and derive the corresponding conditions. In contrast, Lindner and Meyer (2003) study the extremal behaviour of EGARCH processes and prove the extrema of $\ln(\sigma_t^2)$ were light-tailed (in the Weibul domain of attraction), which includes distributions with exponential tails like the Gaussian distribution, and consequently that $\ln(\sigma_t^2)$ would not show cluster behaviour. A further downside of the EGARCH specification for us is that it depends on more parameters, the estimation of which could be more problematic at low frequencies like monthly data, and that the parameter uncertainty problem could be exacerbated by the exponentiation of the variance in EGARCH specification, which could be especially problematic given the use of individual securities as assets in the cross-sectional tests.

⁷⁵ The GARCH model ensures that by posing non-negative coefficients, but in practice in some situation an unconstrained optimisation suggests that negative parameter estimates result in better in-sample fit. EGARCH model would yield positive volatility even if some of the parameters are negative.

⁷⁶ Let M_n be the maximum of a sample of n draws from some distribution $P(X \leq x) = F(x)$. As the sample grows large $n \rightarrow \infty$, the distribution of maxima becomes degenerate as $P(M_n \leq x) = [F(x)]^n$. However, the Fisher-Tippett theorem proves that if there exist normalising series of constants a_n, b_n such that the distribution of the normalised (scaled) maxima $P\left(\frac{M_n - a_n}{b_n} \leq x\right)$ is non-degenerate, then the limiting distribution of extremes can be either a Gumbel distribution (for light-tailed distributions), a Weibul distribution (for distributions with finite support), or a Frechet distribution (for medium and heavy-tailed distributions) (Embrechts et al. (2003)). If a distribution belongs to a Frechet domain of attraction, then the probability of exceeding a certain threshold $P(X > x)$ would have the form $x^{-\alpha}L(x)$, where α is a positive constant and $L(x)$ is some slowly varying function, i.e. $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \forall t > 0$, and so the tail of the distribution decays approximately like Pareto distribution (Gnedenko (1943)). A number of known distributions do not belong to any of these domains of attraction (i.e. they are not max-stable); two important examples are the Gaussian and the exponential distribution.

⁷⁷ The standard tail index results from the previous note were derived under assumption of independent draws, which is clearly not the case with GARCH models, and so the θ referred here is the tail index of the extremes of the volatility series, $\{\sigma_i\}$. The link between the extremes of the process $\{X_i\}$ and the related process $\{\sigma_i\}$ are given by Theorem 4.1 on p. 1437 in Mikosch and Starica (2000)

Overall, each approach has its advantages and disadvantages. The EGARCH removes the difficulties associated with corner solutions to the GARCH optimisation and allows for an asymmetric volatility response to shocks, but it also limits the ability of the model to produce volatility clustering and at monthly frequency could increase parameter uncertainty risk on account of the larger number of parameters to be estimated. Tests with IGARCH and GARCH suggested that IGARCH model did not result in more stable estimates but the accuracy of GARCH was somewhat better. Together with the counter-factual assumption of permanent shocks on variance, we opted for GARCH(1,1) as the baseline model of volatility in this study. We impose the stationarity assumption and specifically constrain the domain for the model parameters to $\alpha + \beta < 1$ and $\alpha, \beta > 0$.

The distribution of the innovations – ν in the EGARCH model (6)-(7) – is another significant choice. Often in related studies ν is assumed to be distributed either as standard normal distribution⁷⁸ or as generalised error distribution.⁷⁹ Both Fu (2009) and Spiegel and Wang (2005) use normal distribution for ν in their EGARCH specifications. Furthermore, Fu allows variable lags for p and q between 1 and 3; among the nine resulting models for each stock and date he selects the one with highest value of Akaike Information Criterion (AIC).⁸⁰ In the GARCH specification the choice of distribution might be a more crucial choice because by design that model reacts symmetrically to innovations and the imposition of a symmetric distribution like the standard normal in the presence of asymmetric innovations⁸¹ might impact the predictive performance of the model. Therefore, we opted for a non-symmetric distribution of shocks. There are a couple of different specifications of the Skew Generalised Error Distribution (SGED). In our study we implemented

⁷⁸ The standard normal (Gaussian) distribution with zero mean and unit variance has density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

⁷⁹ The Generalised Error Distribution has the following three-parameter density:

$$f(x) = \frac{\kappa e^{-\frac{1}{2}\left|\frac{x-\alpha}{\beta}\right|^\kappa}}{2^{1+1/\kappa}\beta\Gamma(1/\kappa)},$$

where α is location parameter, β is scale parameter, and κ is shape parameter. This specification needs to be additionally standardised to zero mean and unit variance. When $\kappa = 2$, the distribution reduces to normal.

⁸⁰ The Akaike's Information Criterion (AIC) is an heuristic derived from information theory that compares competing models in terms of their likelihood (L) on a specific data set and the number of parameters (k), with $AIC = 2k - 2\ln(L)$.

⁸¹ For the full sample skewness of idiosyncratic shocks (arithmetic returns) was positive at about 2.3.

$$f(z) = \frac{\kappa \exp\left\{-\frac{1}{2} \sqrt{\frac{2^{-2/\kappa} \Gamma(1/\kappa)}{\Gamma(3/\kappa)}} z\right\}}{\sqrt{\frac{2^{-2/\kappa} \Gamma(1/\kappa)}{\Gamma(3/\kappa)}} 2^{1+1/\kappa} \Gamma(1/\kappa)} \quad (8)$$

$$f(z|\xi) = \frac{2}{\xi + \xi^{-1}} [f(\xi z)H(-z) + f(\xi^{-1}z)H(z)], \quad (9)$$

where $H(z)$ is the Heaviside step function⁸², ξ is the parameter controlling skewness, and κ is the shape parameter (when κ increases, the distribution gets flatter).

Additional distributions for the innovations were also considered, in particular the Negative Inverse Gaussian, Skew Hyperbolic distribution, Generalised Hyperbolic Skew-Student distribution, and Skew Student- t distribution. We found that SGED tended to produce more robust and accurate forecasts. An intuitive reason could be that the SGED has lighter (exponential) tails; consequently, a strong innovation was less likely compared to the fatter-tailed alternatives, which resulted in quicker adjustment of volatility forecasts upon stronger shock while nevertheless allowing medium tails⁸³ and asymmetric shocks.

Wrapping up the discussion, we use GARCH-family as one of our selected models because it is used in many volatility studies, and by some of the related studies that lend support to the research hypothesis. We chose GARCH(1,1) for one-step forecasts in view of the evidence of its good empirical performance, the scarcity of data at monthly frequency, and the intuitive interpretation of its parameters. We compensate some of the loss of flexibility relative to the EGARCH(p,q) model with varying order by employing a more flexible distribution for the error term (the SGED distribution).

The simple random walk without drift has the form

$$h_t = h_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2),$$

⁸² For practical purposes,

$$H(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z < 0 \end{cases}$$

⁸³ It might be appropriate to note that there are at least three different approaches to discussing tail fatness (heaviness): one approach is to consider the kurtosis of the distribution and to label distributions with kurtosis over 3 as heavy-tailed (3 is the kurtosis of the Gaussian distribution); another approach is based on the previously mentioned concept of max-domains of attraction, with distributions in the Frechet domain being labelled as fat-tailed; finally, heavy-tails could also describe sub exponential distributions, for which the distribution of a sum of sub exponential random variables being asymptotically equal to the distribution of the maximum of the individual variables (i.e. the probability that sum would exceed some threshold is asymptotically distributed as the probability that any single summand would exceed that threshold). The concept of fat tails is discussed in Haas and Pigorsch (2009); in that connection one should note that for example GARCH(1,1) with Normal innovations produces fat-tailed stochastic process in the sense that its kurtosis is over 3. Thus light tails of the shocks do not necessarily mean light tails of the fitted series.

where h_t denotes the stochastic process and ϵ_t is the random shock at time t , which is assumed to be independently and identically distributed (i.i.d.) with mean 0. A recursive substitution shows that $h_t = h_0 + \sum_{i=1}^t \epsilon_i$ and that past shocks never die out but have permanent effect on the level of h_t . If the volatility process is a random walk, then the expected value of h_t conditional on the information at time $(t - 1)$ is just h_{t-1} , which is the known last realisation of the process and reflects all past shocks $\epsilon_i, i = 1 \dots (t - 1)$, i.e. $\mathbb{E}_{t-1} h_t = \mathbb{E}_{t-1} h_{t-1} + \mathbb{E}_{t-1} \epsilon_t = h_{t-1}$. Since the impact of shocks never fades, the variance of the process increases unboundedly as t increases.

Therefore, if volatilities follow random walk, then we could use the volatility in the last month as the expectation of the volatility for the next month. This approach of using $(t - 1)$ volatility h_{t-1} in place of expected volatility $\mathbb{E}_{t-1} h_t$ is employed by Ang et al. (2009, 2006). Even if volatilities are stationary, this does not invalidate entirely the approach of Ang et al.⁸⁴ because persistence of volatilities is a well documented phenomenon. For example, Engle and Patton (2001) recognise volatility persistence as one of the stylised facts of asset returns. Similarly, it is consistent with the approach of Integrated GARCH (Nelson, 1990), which assumes that $\alpha + \beta = 1$ in equation (4). Therefore, if monthly volatility is sufficiently persistent, or if the estimated parameters values of the GARCH(1,1) model are such that $\alpha + \beta$ is sufficiently close to 1, then the approach of Ang et al could result in superior forecasts for h_t ⁸⁵. Therefore, h_{t-1} cannot be discarded as a predictor of $\mathbb{E}_{t-1} h_t$ even in the face of the evidence rejecting the unit root of volatilities.

We include this measure in our study also because it is the one that produced the prediction of a negative correlation between risk and returns, and therefore it is crucial that the evidence is revisited and the finding is confirmed or rejected.

One of the problems with the methodology of Ang et al was that it did not recognise that in the presence of mean reversion, h_{t-1} is not an unbiased predictor of $\mathbb{E}_{t-1}(h_t)$. Fu suggests that GARCH offers a superior solution on the problem. However, there is a subtle difference between the GARCH model and the approach of Ang et al.: the GARCH model uses squared idiosyncratic residuals as a proxy for the realised volatility, whereas the approach of Ang et al. (2006) estimates the volatility in month $(t - 1)$ from higher-frequency data (daily returns). The approach of estimating lower-frequency volatilities

⁸⁴ In their study of 2006 they never claimed that volatilities followed random walk. Instead, they defined their volatility measure to equal the volatility of the previous month.

⁸⁵ Such superiority could be the result from lower noise in the volatility estimates from daily data compared to the squared residuals used in the standard GARCH(1,1).

from higher-frequency data is theoretically justified by Merton (1980), who demonstrates that volatility could be estimated with arbitrary precision using sufficiently high-frequency data. Therefore, the estimation of monthly volatility from daily data could be a completely sound approach and in fact could outperform squared monthly return as a proxy.

Therefore, instead of abandoning the history of monthly volatilities estimated from daily data altogether in favour of GARCH, it might be more justified to attempt to forecast $\mathbb{E}_{t-1}(h_t)$ from the available history. One such approach was pursued by Huang, Liu, Rhee and Zhang (2012) who use ARIMA model fitted on the series of $h_{t-i}, i = 1, \dots, 24$ (i.e. a rolling window design using the last 24 months of data) in order to forecast $\mathbb{E}_{t-1}(h_t)$.

One representation of the Autoregressive Moving Average (ARMA) process with lags 1, ARMA (1,1), is as follows:

$$h_t - m = \phi(h_{t-1} - m) + \theta\varepsilon_{t-1} + \varepsilon_t,$$

where m , ϕ and θ are the parameters of mode. Parameter m is also called the mean-reverting level of volatility because when $\phi \in (0,1)$, the forecasted values of h would converge to m , given enough time. In general, the values of $\phi, \theta \in (-1,1)$ ensure that the ARMA process is invertible.⁸⁶ The unconditional expected value of the process equals m , $\mathbb{E}(h_t) = m$, and ϕ controls the speed of mean-reversion. An alternative parametrisation for the process uses $\mu = m(1 - \phi)$ instead of m , giving the alternative form $h_t = \mu + \phi h_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t$.

Thus a principal difference between using the ARMA(1,1) and GARCH(1,1) in the present context is the proxy they use for the realised volatilities in past months. Indeed, substituting $v_t^2 = \varepsilon_t^2 - \sigma_t^2$ in $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ we obtain $\varepsilon_t^2 - v_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta(\varepsilon_{t-1}^2 - v_{t-1}^2)$, which after re-arrangement yields

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 + v_t^2 - \beta v_{t-1}^2,$$

which is an ARMA(1,1) process in squared shocks (idiosyncratic returns).

Therefore, we include ARMA(1,1) in our study in order to allow the controversial measure of Ang et al. (2006) to be refined in a forward-looking manner, so as to explore how the introduction of forward-looking expectations amends the conclusions reached by Ang et al. (2006) and Ang et al. (2009).

A principal limitation of historical volatilities is that they incorporate only historical

⁸⁶ The formulation of the ARMA process means that all observed values of h_t are function of the history of unobserved errors. If the process is invertible, then these errors can be represented as weighted sums of the observed realisations.

information and are not, in fact, truly forward-looking. It could be expected that the use of implied volatilities should be producing more a conclusive test of the models of Levy (1978) and Merton (1987). However, derivatives are written only on some of the larger stocks, so that approach could not be pursued.

3.2.5.4. Comparison of volatility forecasts

There are many methods that can be used to evaluate alternative volatility forecasts. These differ in terms of the forecasting scheme, the measure of forecast quality, the estimator of the true latent volatility, and the approach for testing the significance of differences of accuracy among the alternative forecasts.

Concerning the forecasting schemes, Violante and Laurent (2012) highlight three principal schemes: fixed, rolling, and recursive. In the fixed scheme, parameters of the model are estimated using a set of fixed length, and then all future forecasts are made without re-estimation of parameters. The rolling scheme is implemented using a rolling window of fixed length. Finally, a recursive scheme is implemented using all past information to make each forecast, with parameters being re-estimated for each period. In our situation we are in fact unable to pick one of these, because the different forecasts used in the studies are essentially estimated with different approaches. Thus, the OLS estimator and the estimator of Ang et al. (2006) are estimated using rolling windows (in the latter case – of length just one), while GARCH and ARMA forecast are based on expanding windows (recursive scheme). In general, Violante and Laurent (2012) note that the fixed and the rolling schemes have certain advantages over the rolling one when comparing nested models. Firstly, the fixed scheme could be useful in situations where parameter estimation is difficult. Secondly, the rolling scheme could accommodate situations where the estimated parameters change over time. Finally, there are difficulties in implementing statistical tests in the recursive scheme due to the complex asymptotic distribution of the test statistics as sample size grows with each recursive estimation. In our case, however, the use of non-linear models requires larger samples, and therefore the rolling scheme cannot be implemented for some of the volatility estimators (GARCH and ARMA). Furthermore, such an implementation, if possible, would render the produced forecasts incompatible with those in other studies.

The second problem which we face in the tests is that variances are latent (unobservable), and therefore need to be estimated. Violante and Laurent (2012) note that there are three typical estimators: the squared returns, the realised volatility, and the realised kernel. The squared returns are an unbiased estimator when expected return is zero, as from

probability theory we have $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$, so when $\mathbb{E}X = 0$, the variance of X equals the expected value of X^2 . The realised variance can also be calculated from the returns of higher frequency. For example, the monthly variance can be estimated from daily returns as $\text{var}(X_M) = \sum_{d \in M} X_d^2$. Even with unchanged daily variance, monthly variances will change depending on the number of days in the month. In cases where such changes were a concern, we averaged volatilities by the number of days in month to calculate the daily variance, and then scaled to a standardised month of 21 trading days. The last estimator of true volatility – the realised kernel, developed by Barndorff-Nielsen et al. (2008) – is used mostly in high-frequency trading where microstructure noise is a significant problem. The method essentially employs kernel smoothing of the higher-frequency volatilities in order to smooth out microstructure noise. The method is rarely used for estimation of monthly volatilities. Moreover, the parameters of the smoother would need to be tuned, which would increase the complexity of estimates with unclear benefit. Instead, studies like Bali and Cakici (2008) and Spiegel and Wang (2005) filter the daily volatility series through a GARCH model which smooths out noise from daily returns and then calculate the monthly realised volatility from the daily filter volatilities. We follow that approach as well.

We consider three estimators of the true variance. The first is the traditional but noisy squared idiosyncratic shock, i.e. $h_t = IRet_t^2$. In order to obtain a more accurate estimate we also estimate an EGARCH(1,1) model using all available *daily* data. The in-sample estimates for idiosyncratic variance are averaged by months (we require estimated variances for at least 10 days in each month). Thus $h_t = \frac{1}{\tau_t} \sum_{i=1}^{\tau_t} \sigma_i^2$, where τ_t is the number of days in month t . EGARCH(p , q) is defined by the following volatility equation:

$$\begin{aligned} \log(\sigma_t^2) = & \omega + \sum_{k=1}^p \beta_k \log(\sigma_{t-k}^2) + \\ & \sum_{m=1}^q \alpha_m \left\{ \theta \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) + \gamma \left[\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \mathbb{E} \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) \right] \right\}. \end{aligned} \quad (10)$$

The distribution of $\varepsilon_{t-1}/\sigma_{t-1}$ is assumed to be standard Gaussian, so $\mathbb{E}(\varepsilon_{t-1}/\sigma_{t-1}) = (2/\pi)^{1/2}$. The specification allows an asymmetric response of volatility, a phenomenon observed in equity index returns. When $\gamma < 0$ in the above model, a negative shock ε_{t-1} increases volatility more than a positive shock of equal magnitude would have done. The mean equation for the volatility model is again the Fama–French–Carhart four-factor model estimated using OLS. Finally, as a control version we also filter monthly volatilities from the GARCH(1,1) model with SGED innovations, estimated on all available monthly idiosyncratic returns.

There are two principal methodologies for comparison of volatility forecast that were employed in existing studies: the loss function approach, and the Mincer-Zarnowitz regressions. The former (loss function approach) is useful in comparison of alternative volatility estimators estimated from identical target frequencies, i.e. not requiring scaling from one frequency to another. An example could be alternative GARCH specifications, e.g. the implications of different distributions of errors on the produced GARCH(1,1) forecasts. Mincer and Zarnowitz (1969) regressions, however, can be employed to compare forecasts that come from different frequencies, in order to avoid the problem of data scaling.

The loss function approach compares forecasts based on cost functions that assign weight on the distance between the volatility forecast and the estimate of the true (*ex post*) volatility. Spiegel and Wang (2005) utilise that approach to compare the predictive performance of EGARCH and OLS idiosyncratic variance estimators.⁸⁷ Hansen and Lunde (2005) analyse the performance of various cost functions and recommend two of the options: the mean squared error ('MSE') and the quasi-likelihood cost function ('QLIKE'):

$$MSE_i = \frac{1}{T} \sum_{t=1}^T (h_{i,t} - \hat{\sigma}_{i,t}^2)^2$$

$$QLIKE_i = \frac{1}{T} \sum_{t=1}^T \left(\ln(\hat{\sigma}_{i,t}^2) + \frac{h_{i,t}}{\hat{\sigma}_{i,t}^2} \right)^2,$$

where $h_{i,t}$ is the observed true (*ex post*) variance and $\hat{\sigma}_{i,t}^2$ is the *ex ante* conditional expected variance. In the baseline case the true variance is inferred from squared returns, $h_{i,t} = r_{i,t}^2$. Another possible measure used in the field is the mean absolute error ('MAE')⁸⁸:

$$MAE_i = \frac{1}{T} \sum_{t=1}^T |h_{i,t} - \hat{\sigma}_{i,t}^2|.$$

The estimated values for the cost functions depend on the noisy estimate for the true variance (the squared idiosyncratic shocks, $h_{i,t} = r_{i,t}^2$). Differences in MSE or QLIKE do not necessarily imply that one forecasting method outperforms another. Various parametric and non-parametric tests can be employed to test forecast performance, e.g. Diebold and Mariano (1995), Meese and Rogoff (1988), or Granger and Newbold (1986) tests. Inference based on those tests relies on an asymptotic distribution of the test statistics, which could be

⁸⁷ Spiegel and Wang (*ibid.*) report that EGARCH halves the mean absolute error of the estimated variance compared to the variance of the residuals (OLS) from the mean equation, although we were unable to confirm such an improvement in the full sample.

⁸⁸ An advantage of the MAE metric is that it does not assume that the second moment (variance) of the tested series exists (point on p. 12 in Meese and Rogoff (1983)), which would be the case if variances follow random walk. We do not think this would be the case here, as the unit root hypothesis is rejected by Fu (2009) for 90% of the series using the Dickey-Fuller test, and our calculations yield very similar results.

unreasonable in cases where we have few observations. Therefore, we follow the suggestion of Diebold and Mariano (1995) to employ the sign test or the Wilcoxon's signed rank test when a comparison of loss functions or other measures of prediction accuracy is required.

The sign test is based on the observation that if the difference between two estimators is not systematic, then the median difference between the two series would zero. One could then test if that is the case by performing binomial test with probability equal to 0.5. We can calculate those differences, for example, using the differences between squared or absolute values:

$$d_{i,t} = (h_{i,t} - \hat{\sigma}_{1,i,t}^2)^2 - (h_{i,t} - \hat{\sigma}_{2,i,t}^2)^2,$$

or

$$d_{i,t} = |h_{i,t} - \hat{\sigma}_{1,i,t}^2| - |h_{i,t} - \hat{\sigma}_{2,i,t}^2|.$$

Therefore, for a given security, the median of the two series above should equal zero.⁸⁹ The same idea could be applied also if we compare loss functions (MSE, QLIKE, MAE) across securities by calculating the differences $d_i = \text{LF}(h_i, \hat{\sigma}_{1,i}^2) - \text{LF}(h_i, \hat{\sigma}_{2,i}^2)$, where $\text{LF}(h_i, \hat{\sigma}_{k,i}^2), k = 1,2$ is the loss function for security i calculated using its available history.

The test statistic then becomes simply

$$S_{sr} = \sum_{i=1}^N 1_{d_i>0} \sim \text{Binom}(N, 1/2),$$

where $1_{d_i>0}$ denotes the indicator function. In large samples the normal approximation could be employed.⁹⁰

The same idea of differences could be employed to calculate the Wilcoxon's signed rank test statistic, which equals the sum of the ranks of the absolute differences for the cases where the difference is positive:⁹¹

$$S_w = \sum_{i=1}^N 1_{d_i>0} \text{Rank}(|d_i|).$$

Besides the tabulated critical values for the test, normalised approximation could also be

⁸⁹ This imposes the implicit assumption of pairwise complete forecasts, so that cases where an estimate from one estimator was available while for the other series was unavailable would not participate in the joint test.

⁹⁰ i.e. for N large, $\frac{S_{sr}-0.5N}{\sqrt{0.25N}} \sim N(0,1)$.

⁹¹ Since $\sum_{i=1}^N i = \frac{N(N+1)}{2}$, we expect that the sum of the ranks of the positive differences would be one half of the amount.

employed.⁹²

Another approach to assess goodness of fit relies on estimating Mincer and Zarnowitz (1969) regressions of the type

$$h_{i,t} = a_i + b_i \hat{\sigma}_{i,t}^2 + e_{i,t}, \quad (11)$$

where $h_{i,t}$ is the estimate for the true (realised) idiosyncratic variance for security i at month t , and $\hat{\sigma}_{i,t}^2$ is the *ex ante* estimate for that variance. If one model predicts volatilities correctly, we should expect to fail to reject the joint null hypothesis that $a_i = 0, b_i = 1$, so in practice we test whether $(h_{i,t} - \hat{\sigma}_{i,t}^2)$ has mean 0. When applied to GARCH models, a common finding is that the Mincer-Zarnowitz regressions have quite low explanatory power as measured by their R^2 's. Andersen and Bollerslev (1998) prove that the reason for the low R^2 's is not a failure of the GARCH models, but rather the fact that squared returns are a noisy estimator of volatility. In particular they show that the R^2 's for GARCH(1,1) models are bounded above by $1/\kappa$, where κ is the kurtosis of the underlying noise distribution; therefore, for normally-distributed noise, $R^2 < 1/3$, and for heavy-tailed distributions the bound is even lower.

Running the Mincer-Zarnowitz regressions in variances is sometimes criticised as placing undue weight on high volatilities episodes, or being affected by the skewness of the volatility distribution. Thus Bali and Cakici (2008) use $\widehat{IV}_{t-1}^{daily}$ and a rolling average of $\widehat{IV}_{t-1}^{daily}$ but instead of running the regression in variance, they run it into volatilities (standard deviations), while Pagan and Schwert (1990) also report the regressions for the logs of variances. Therefore we address such concerns by employing all three regression models, in order to examine the predictive performance of volatility forecasts:

$$\begin{aligned} h_{i,t} &= a_i + b_i \hat{\sigma}_{i,t}^2 + e_{i,t}, \\ \sqrt{h_{i,t}} &= a_i + b_i \hat{\sigma}_{i,t} + e_{i,t}, \\ \ln(h_{i,t}^2) &= a_i + b_i \ln(\hat{\sigma}_{i,t}^2) + e_{i,t}, \end{aligned}$$

where h_t is the realised variance and $\hat{\sigma}^2$ is the forecast variance.

The usual OLS estimator is sensitive to extremes. In order to mitigate possible concerns that differences in the average R^2 statistic are driven by few larger outliers, we also

⁹² i.e. for N large, $\frac{S_w^{-\frac{N(N+1)}{2}}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \sim N(0,1)$.

estimate the parameters of the three models specified above using quantile regression⁹³. Unlike the OLS regression, which minimised the sum of squared errors and estimated the conditional mean of the explained variable, the quantile regression estimates some conditional quantile τ of the explained variable. In this case we aim to predict the conditional median ($\tau = 0.5$), which we accomplish by minimising the sum of the absolute differences, i.e. $V_\tau = \min_\beta \sum_i \rho_\tau(y_i - x_{it}\beta)$, where $\rho_\tau(u) = u(\tau - I_{u < 0})$. In the case of forecasting of the median ($\tau = 0.5$), the problem reduces to $V_{0.5} = \min_\beta \sum_i |y_i - x_{it}\beta|$. An obvious constraint would be that R^2 is not directly applicable to this model. Instead, following Koenker and Machado (1999) we compute its analogue for quantile regression – R_τ^1 , which is defined as $R_\tau^1 = 1 - V_\tau/V_\tau^c$, where V_τ is the optimal cost function for the evaluated model, and V_τ^c is the optimal cost function for a constrained model where all predictors (other than the intercept) are constrained to equal zero. In this way R_τ^1 measures how much the model improves the prediction of the conditional median relative to the model of constant median. Thus, R_τ^1 is a direct analogue of R^2 in the domain of quantile regression, and we use it to evaluate how the competing predictors of the volatility perform.

Overall, we compare volatility forecasts in terms of Mincer and Zarnowitz (1969) regressions run in variances, standard deviations, and logarithms. Forecast accuracy is compared using Wilcoxon’s signed rank test for the pairwise R^2 from each predictive regression ran on security by security basis.

3.2.5.5. Assessing the correlation between expected idiosyncratic volatility and returns

There are at least three types of tests of asset pricing models which are based on somewhat different formulations of the asset prices. The stochastic discount factor (SDF) approach proposes that prices equal present values of the end-of-period prices and dividends, discounted by SDF (m_{t+1}), so that $p_t = \mathbb{E}_t(m_{t+1}x_{t+1})$ (the pricing equation), and hence $\mathbb{E}_t(m_{t+1}(1 + R_{t+1}) - 1) = 0$, where $R_{t+1} = \frac{x_{t+1}}{p_t} - 1 = \frac{p_{t+1} + d_{t+1}}{p_t} - 1$ denotes the return on the asset, p_{t+1} is the future price, and d_{t+1} is the future dividend. Thus, in this framework, the valuation of prices requires estimation of the stochastic discount factor, m_{t+1} . There could be different ways to impose structure on the SDF. One approach, e.g. used in the Consumption CAPM (CCAPM) and the Lucas exchange economy, assumes utility function of

⁹³ See Koenker and Gilbert Bassett (1978)

specific type to deduce the structure of the SDF. For example, assuming additive lifetime utility function in consumption of the form $U = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_0[u(c_t)]$, so that the utility is time-separable and lifetime utility equals the present value of the utility of the future consumption stream, CCAPM concludes that $m_{t+1} = \frac{\beta u'(c_{t+1})}{u'(c_t)}$, where β is the intertemporal discount factor, and $u'(c_t)$ is the marginal utility of consumption. Another approach assumes that m_{t+1} can be approximated linearly by a set of factors, so that $m_{t+1} = a + b' f_{t+1}$, where a is parameter vector, b is parameter matrix, and f_{t+1} is the vector of factors.

The beta representation of the pricing equation follows from the transformation $\mathbb{E}_t(m_{t+1}(1 + R_{t+1})) = 1$, so that $\mathbb{E}_t(m_{t+1})\mathbb{E}_t(1 + R_{t+1}) + \text{cov}(m_{t+1}, (1 + R_{t+1})) = 1$. Dividing by $\mathbb{E}_t(m_{t+1}) > 0$, the approach yields $\mathbb{E}_t(R_{t+1}) = \frac{1}{\mathbb{E}_t(m_{t+1})} + \frac{\text{cov}(m_{t+1}, R_{t+1})}{\text{var}(m_{t+1})} \left(-\frac{\text{var}(m_{t+1})}{\mathbb{E}_t(m_{t+1})} \right)$. The term $\beta = \frac{\text{cov}(m_{t+1}, R_{t+1})}{\text{var}(m_{t+1})}$ is the slope of the regression of returns on the SDF (and thus by the linear approximation of SDF in terms of factors, betas are related to factors), while $\lambda = -\frac{\text{var}(m_{t+1})}{\mathbb{E}_t(m_{t+1})}$ is the price of risk. Extensive treatment of these derivations and the relationships between SDF, factor models, and beta representations is available in Gospodinov and Robotti (2013), as well as Chapters 5 and 6 in Cochrane (2005). For the purposes of the present discussion it is important to note that these relationships can be tested in alternative forms, and these forms are related by the alternative specifications above. In particular, time-series tests can be implemented by regressing excess returns on the factor realizations, i.e. $(R_t - R) = \alpha + \beta_1 f_{1,t} + \beta_2 f_{2,t} + \dots + \beta_k f_{k,t}$. This approach is essentially employed by us to split total returns into systematic and idiosyncratic returns using three (Fama–French) or four (Fama–French–Carhart) factors. However, that approach is not useful to perform the test whether idiosyncratic volatility predicts returns because volatilities are not factors; furthermore, as we shall see in the chapter on results, some of the forecasts are non-stationary, which also prevents running the time-series regressions in volatilities.

Another approach that is much more widely used in finance tests asset pricing models using the Generalised Method of Moments (GMM). There is significant body of literature on the use of those methods, but the principal idea is to find parameter values that would minimize the distances between the moments of the left-hand side and the right-hand side of the pricing equation $p_t = \mathbb{E}_t(m_{t+1} x_{t+1})$. The mean pricing error in a sample of size T then has the form $g_T = \frac{1}{T} \sum_{t=1}^T (m_{t+1} x_{t+1} - p_t) = \mathbb{E}(m_{t+1} x_{t+1} - p_t)$, and the GMM parameters are estimated by minimizing the weighted errors, $g_T' W g_T$, where W is the matrix of weights

(in case of equal weights, $W = I$, where I is the identity matrix). The method allows the estimation of general non-linear, arbitrage-free asset pricing models. Nevertheless, it also has the disadvantage that the choice of base instruments or parameterization could result in unstable estimates. Compounded with the greater complexity of estimation, the method is not often used in the related literature.⁹⁴

The third method used in literature is based on cross-sectional regressions, and it is the method used in this study. A principal advantage over the time-series estimation is that the cross-sectional methodology can easily accommodate the addition of characteristics or factors that are not returns.

The original methodology for the cross-sectional tests is proposed by Fama and MacBeth (1973). They treat the coefficients from the cross-section regressions as random variables and employ the t -test to examine the statistical significance of the different characteristics. Thus, Fama and Macbeth estimate the cross-sectional regression separately for each month, instead of averaging returns and betas in the testing period. The cross-sectional regressions has the form:

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t}\beta_{i,t} + \gamma_{2,t}IVOL_{i,t} + \dots + \varepsilon_{i,t}, \quad (12)$$

where $IVOL$ is some of the measures of idiosyncratic volatility employed in this study, and $Beta$ is the beta for security i .⁹⁵ Many more characteristics could be added to the regression as required for hypothesis testing, e.g. size, liquidity, lagged return, momentum, etc. The significance of the hypothesised factor loadings ($\hat{\gamma}_i$) could be tested using the t -statistic:

$$t(\hat{\gamma}_i) = \frac{\bar{\gamma}_i}{s(\hat{\gamma}_i)/\sqrt{n}},$$

where n is the number of cross-sectional regressions, $s(\hat{\gamma}_i)$ is the standard error of the regression coefficient, and $\bar{\gamma}_i$ is the mean value of the respective factor loadings, averaged across all monthly cross-sectional regressions. Essentially, the test treats the coefficients of the cross-sectional regression as independent random variables and performs a standard test whether the mean realisation is statistically different from zero. Newey and West (1987) propose a refinement of the t -test that allows to incorporate the autocorrelation of the coefficients from the individual cross-sectional regressions. Following the prevailing practice in the related literature, in this thesis we report the Newey–West standard errors with four lags.

⁹⁴ A notable exception is Khovansky and Zhylyevskyy (2013)

⁹⁵ We discuss the estimation of beta later in this chapter.

3.2.5.6. Securities as assets

The literature on testing whether higher idiosyncratic returns are associated with higher return usually employs two strategies: cross-sectional regressions, and portfolio formation. In this study we employ the former strategy. In our view there are two advantages of that strategy over the latter one: firstly, it allows tests based on individual securities as assets (as opposed to portfolios as assets), which is suggested to result in superior efficiency of the performed tests⁹⁶. Secondly, portfolios are created in two dimensions, one of which was the hypothesised dimension (idiosyncratic risk) and the other was the control dimension (beta, size, liquidity, etc). In the tests of idiosyncratic risk, however, there are at least four significantly correlated characteristics, each with its own set of theoretical justifications. Those characteristics are: beta with market, size, liquidity, and idiosyncratic risk. Larger companies empirically tend to have lower beta, higher liquidity, and lower idiosyncratic risk. The tests based on portfolio formation aim to demonstrate that higher-idiosyncratic-risk portfolios would earn higher expected return irrespective of the control variable used for the other dimension. Nevertheless, one could not eliminate the risk that the observed positive returns were due to some of the other two non-controlled characteristics. This is supported for example by the study of Fan et al. (2015) who demonstrate a significant link between market anomalies and idiosyncratic risk. Therefore, in this study we opt to test the link between idiosyncratic risk and return principally in the cross-section of returns rather than through portfolio formation.

The cross-sectional regressions, however, can also be implemented using portfolio returns and volatilities as assets, rather than individual securities. In this study we have chosen the second option (individual securities as assets). There are a couple of motivations for this choice. Firstly, the use of portfolios as assets could result in loss of efficiency of the tests. The portfolio strategy was originally conceived to address errors in variables in the estimation of betas, which can be estimated with significant errors. However, most controls can be estimated with high precision, including size, price/book values, return momentum, lagged return, and idiosyncratic volatility. Therefore, Fama and French (1992) argue that the use of portfolios is not justified, apart from the case of beta. The conjecture that use of portfolios leads to loss of efficiency is confirmed by Ang et al. (2010), who point out that even betas should not be averaged by portfolios. Secondly, as pointed out by Levy (2012), the Fama and MacBeth (1973) methodology “employs portfolios rather than individual assets; therefore, it

⁹⁶ Ang et al. (2010)

has the advantage of minimising the measurement errors in beta and the disadvantage of not testing asset pricing of individual assets. Thus, in the case of supporting the CAPM, one cannot generalise it to individual risky assets.” (p. 200). In our study we aim to evaluate whether idiosyncratic volatilities are priced, and therefore averaging volatilities by portfolios would likely result in loss of efficiency. Nevertheless, we also run robustness checks using portfolios as assets to confirm our principal findings.

3.3. Control variables

The statistical tests for the cross-sectional correlation between idiosyncratic risk and returns necessitate the identification of a set of variables that are hypothesized to explain the stock returns both as time series and as cross-section. Therefore, we need to identify firstly the variables that we shall employ to split the time series of actual returns into systematic and idiosyncratic components. Then we shall identify control variables for the cross-sectional regressions. At first sight the two lists could be hypothesised to be identical, i.e. we use a list of factors to split returns into systematic and cross-sectional components, and then use the loading on those factors in the cross-sectional regression, e.g. estimate asset betas with market excess returns, and then use the betas as explanatory variables in the cross-section. In such a setting, the time series tests and the cross-sectional tests would be testing the same hypothesis, so arguably there would not even be any need for two steps. However, we noted previously that the time-series regressions require the use of factors as explanatory variables, so some of the variables that we aim to use in our test, including idiosyncratic volatility, cannot be used in the time series regression. Therefore, we consider a list of factors in order to split returns into systematic and non-systematic (idiosyncratic) components, and then we use the beta together with characteristics that measure the individual asset’s exposure to the hypothesised factors, as well as other characteristics considered relevant.

3.3.1. Splitting total return into systematic and idiosyncratic returns

The Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965b) predicts that there is a single factor that explains the mean excess returns of individual stocks, and that factor is the excess return on the market portfolio, i.e.

$$\mu_{i,t} = r_f + \beta_i(\mu_{m,t} - r_f).$$

The deviation of the actual return from the expected value is due to idiosyncratic factors, $\varepsilon_{i,t} = r_{i,t} - \mu_{i,t}$, but those idiosyncratic innovations are diversified away and are not priced by the investors. The tests of CAPM are challenged by Roll (1977), who shows that the market portfolio to which CAPM is referring is the portfolio of all assets in the economy, including human capital, real estate, privately held businesses, overseas assets. That portfolio is not observable, and its replacement by equity index introduces an error-in-variable problem, which could account for observed anomalies detected by some of the empirical tests.

Fama and French (1992, 1993, 1996) isolate three factors that explain a significant part of cross-sectional returns: market excess returns (from the CAPM), a size factor and a value factor. The size (SMB) and value factors (HML) are estimated by splitting the stocks into six sub-portfolios in two dimensions by ranking the portfolios in terms of size (three groups) and in terms of book-to-market value (two groups). The realisation of the SMB factor is estimated by considering a portfolio that is long in small shares and short in large shares. Similarly, the realisation of the HML factor is estimated by using the returns to a no-cost portfolio that is long in value stocks (those with high book-to-market ratio) and short in growth shares (those with low book-to-market ratio). The approach of Fama and French is justified in terms of construction of the predictive indices, but it is criticised for not providing theoretical motivations as to why investors should value these factors. The SMB factor is sometimes interpreted as a measure of probability of distress – smaller companies are less diversified and with more limited access to funding, which increases their probability of distress, which is priced by the market. The reasoning behind the value factor, however, is less conclusive.

More recently, Fama and French (2007) investigate the contribution of the sub-components of returns for value and for growth stocks. In the case of value stocks they find that the positive returns stem from increases of price-to-book value ratio, while the amount of equity is broadly stable. In the case of growth companies, returns came from the strong increase of equity that is enough to outpace the corresponding decline of the price-to-book value ratios. They suggest that eventually after the stocks are allocated to the corresponding portfolio, the market causes the differences in performance of value and growth portfolio to blur. In the case of growth stocks this is due to the gradual exhaustion of the high-profit opportunities and strengthening competition, while in the case of value companies it comes on the back of the urge to improve profitability, which for value companies is below that of growth companies. Consequently some of the growth stocks migrate to the value group and some of the value stocks migrate to the growth portfolio.

One stock market anomaly that the Fama–French specification does not account for is stock momentum. Jegadeesh and Titman (1993) document a tendency for portfolio that buys past winners and sells past losers to outperform the market over the sample period 1965 to 1989. To examine that effect, they calculate the return on each share over the past J months. Then they allocate shares in ten decile portfolios ranked by their past performance, and track those portfolios K months ahead. They call ‘winners’ the top decile portfolio (the one that contains 10% of shares with the highest performance in the preceding J months), and ‘losers’ – the bottom decile portfolio (the shares with poorest performance in the past J months). Examining various buy/sell strategies for various values of the past return calculation window J and holding period K , they find that strategies that buy past winners and sell past losers earn significant premium. For example, the strategy which selects stocks based on their performance in the past $J = 6$ months and holds them for $K = 6$ on average earns a compounded excess return of 12.01% p.a. The results are also found to hold when portfolios are created from sub-samples formed on rankings on systemic risk (beta) and size (capitalisation), so momentum effect is not due to systemic risk or size factor. Furthermore, the buy-winners/sell-losers strategy earns abnormal return for horizons up to 36 months after buying, although the last 24 months tend to reverse the gains from the first 12 months. They find that the observed momentum could not be explained by lagged response to common factors but is consistent with lagged response to firm-specific information, with performance systematically strong in the first year following buying, except in the first month. Scowcroft and Sefton (2005) investigate the sources of momentum and decompose it into country momentum, sector (industry) momentum and the residual idiosyncratic momentum. They find that for small-capitalisation companies momentum is primarily idiosyncratic, while for large-capitalisation companies it is mostly industry specific.

Bondt and Thaler (1985) report that portfolios of past ‘losers’⁹⁷ significantly outperform the portfolio of past ‘winners’⁹⁸ and find that the over-performance lasts for up to 36 months after portfolio formation. They attribute the observed differences to market over-reaction that is reversed subsequently, albeit very slowly, as evidenced by the horizon over which the discrepancies persist.

In general, there is no constraint on what factor model is selected for the calculation of

⁹⁷ The top 35 securities/top 50 securities/the top decile with lowest cumulative abnormal (residual) return over the 36 months preceding portfolio formation

⁹⁸ The same number of securities with the highest cumulative abnormal return over the same period as for the identification of past ‘losers’

idiosyncratic return, as long as we remain alert that different models result in different residuals. Thus idiosyncratic residuals need not necessarily reflect pure idiosyncratic risk but could reflect return on omitted factors⁹⁹, e.g. return on human capital¹⁰⁰. The early tests of CAPM define idiosyncratic risk relative to the single-factor CAPM model. Among the recent studies, Malkiel and Xu (2004) report results for idiosyncratic risk from the CAPM (they also report results from the three-factor model). The one-factor specification is also used by Bali and Cakici (2008) in their sorts by idiosyncratic volatility, although performance is assessed based on alphas from the Fama–French specification discussed below.

More recent studies tend to base their calculation of idiosyncratic return on the CAPM extended with the two additional Fama–French factors (this specification we shall abbreviate as ‘FF-3’). The preference for FF-3 in recent finance literature is driven by the good empirical performance of that specification. Most of the reviewed studies on the link between idiosyncratic risk and the cross-section of returns define idiosyncratic risk relative to the FF-3 specification, e.g. Spiegel and Wang (2005), Fu (2009), Brockman et al. (2009), Guo et al. (2014), Fu and Schutte (2010). More recently, Huang, Liu, Rhee and Wu (2012) add momentum (‘MOM’) factor when calculating idiosyncratic returns. This four-factor specification is usually referred as the Fama–French–Carhart (‘FFC’) specification by the name of Carhart (1997), who originally added the momentum factor.

In this study we shall employ the four-factor FFC specification to separate monthly returns into systematic and idiosyncratic components, i.e.:

$$(r_{i,t} - r_{f,t}) = \beta_0 + \beta_1(r_{M,t} - r_{f,t}) + \beta_2r_{SMB,t} + \beta_3r_{HML,t} + \beta_4r_{MOM,t} + \varepsilon_{i,t}, \quad (13)$$

where $(r_{i,t} - r_{f,t})$ is the excess return of asset i over the risk-free rate¹⁰¹, $(r_{M,t} - r_{f,t})$ is the excess return on the market value-weighted portfolio, and $r_{SMB,t}$ and $r_{HML,t}$ are the returns on the Fama–French factors, and $r_{MOM,t}$ is the return on the momentum factors. Note that $\varepsilon_{i,t}$ are the errors of the regression, but are not the idiosyncratic innovations used in this study. The equation above is used to forecast next period systematic returns from the factor

⁹⁹ “Since volatilities, especially idiosyncratic volatilities, are unobservable, most empirical studies estimate them using residuals from fitting a market model. Empirically, however, it is very difficult to interpret the residuals from the CAPM or even a multi-factor model as solely reflecting idiosyncratic risk. One can always argue that these residuals simply represent omitted factors. Therefore, we can only assert that the residuals from a market model measure idiosyncratic risk in the context of that model.” (p. 19 in Malkiel and Xu (2004))

¹⁰⁰ As proposed in Eiling (2013)

¹⁰¹ We use the yield on the three-month constant maturity government bonds as a measure of the risk-free rate.

realisations, and then the difference to the total return would be the idiosyncratic risk.

The choice of FFC over FF-3 is motivated by two considerations. Firstly, the evidence in favour of the momentum factor is now reasonably well established, and we have previously referred to some of the papers documenting that finding. Secondly, as documented by Fan et al. (2015), idiosyncratic volatility significantly correlates with return momentum, and therefore it may be argued that its omission results in idiosyncratic residuals containing momentum information. On the other hand, the impact of one factor specification over another should not be overstated. For example, the reviewed study by Malkiel and Xu uses both CAPM and FF-3 specifications but does not find significant differences in their results. There Malkiel and Xu (2004) commented that “we use idiosyncratic volatility estimates both from a market model and from the above Fama-French three-factor model. Since residual volatility is a second moment, we view this approach as an indirect control for other factors.” (p. 19). Inasmuch as the two-factor model specifications did not provide qualitatively different results, the study can be viewed as offering evidence that the addition of extra factors is not a driver of the obtained results. At any rate the pass-through from omitted factor variable to idiosyncratic volatility would be limited, as pointed out by Malkiel and Xu (2004) in connection with the impact of omitted liquidity factor: “If liquidity is indeed priced, residuals from any asset pricing model that excludes liquidity factor will reflect it. However, since idiosyncratic volatility is a second moment, it can only indirectly capture some of the liquidity effect.”(p. 32)

We recognise that there may be arguments to include even more factors in the time-series regression above. For example, Fama and French (2015) propose a five-factor model that besides size and value, also adds profitability and investment patterns. However, Kan and Zhang (1999b) point out that the inclusion of an irrelevant factor in the model makes the covariance matrix non-invertible, and the asymptotic properties of the two-step regression tests are adversely affected.¹⁰² In view of the risks in adding extra factors to the regressions, the fairly small history used for model estimation, as well as the limited accounting information, we decided against adding more factors in the time-series regressions. We nevertheless include a specific robustness test with statistical factor in order to verify our findings.

¹⁰² Kan and Zhang (1999a) find a similar situation with the use of the generalized method of moments.

3.3.2. Control variables in the cross-section regressions

We now turn to the problem of selection of control variables for the cross-sectional regression. At first glance the most direct approach could seem to use the estimated loadings on the four factors as regressors in the cross-sectional regressions, together with other control variables, in particular – idiosyncratic volatilities. Such an approach would be consistent from a theoretical perspective but is not actually pursued in the literature because of the uncertainty of the estimation of betas. Therefore, betas of individual securities are usually replaced by betas of portfolios;¹⁰³ imposing such a constraint on the value of beta that is different from that estimated from the time-series regression means that the other factor loadings are no longer guaranteed to be unbiased, and therefore necessitates the use of other measures of size, value and momentum, other than the slopes from the time-series regressions.

The CAPM justifies the use of asset's beta with the market as an explanatory variable. However, in view of the measurement error problem in its estimation, we shall follow the process proposed in Fama and French (1992), and we shall form size-beta portfolios, and replace individual betas with those of the portfolio, to which the asset is assigned in the given period. Likewise, the exposures to size and value factors shall be measured more directly in terms of (the logs of) capitalisation for the given security and the market/book value ratios. Exposure to the momentum factor can be estimated using the cumulative return over a six-month period. However, the value of the last return would have a specific significance, because of a mild negative autocorrelation in the data (negative returns are more likely to be followed by a positive return), and therefore the six-month period ends at the penultimate month, rather than at the last month. Finally, we also include a liquidity variable as it tends to be correlated with idiosyncratic volatilities (Spiegel and Wang, 2005) and could be hypothesized to be the cause of the predictive significance of idiosyncratic volatilities.

Additional control variables were also considered, e.g. measures of tail fatness or return skewness, but proved insignificant predictors of the cross-section. In view of the aforementioned risk of confirming spuriously significant premium for insignificant factors, explored by Kan and Zhang (1999b), throughout this study we shall emphasise not only statistical significance, but also the stability of the estimates across difference specifications and subsamples. If a certain variable is a significant predictor, we would expect that its point

¹⁰³ This is considered as inefficient by Ang et al. (2010), but thus far the prevailing practice has been to use portfolio betas, and we follow the other studies in that respect.

estimate remains fairly stable, and changes in its direction can be reconciled with economic intuition.

3.4. Data sources, transformations, and summary statistics

3.4.1. Data sources

In this section we describe the data set that we use for our empirical analysis, as well as the precautions and procedures taken to ensure the quality and integrity of the data set.

The primary data source for the stock prices and the economic time series used in this study is Thompson Reuters Datastream and save for the Fama–French factors, all other data items are sourced from there.

The data set covers all shares (Datastream instrument type = “equities”). Studies in the field, e.g. Campbell et al. (2000), Ang et al. (2006), use stocks listed on NYSE, AMEX and NASDAQ, and we have followed suit and included all of them in our sample. Specifically we require that the included equities have a primary listing on the New York Stock Exchange (henceforth “NYSE”), the NASDAQ Stock Market¹⁰⁴ (henceforth “NASDAQ”), or the NYSE MKT exchange (henceforth “NYSE MKT” or “AMEX”¹⁰⁵). NYSE, NASDAQ and NYSE MKT were respectively the first, the second and the third largest American stock exchanges.

For all securities we require that the market should be the United States of America and the currency of the issue should be the United States Dollar. The limitation of the scope to the United States is intended to facilitate comparison with our reference studies, and also to alleviate problems with country-specific factors of assets returns stemming from country-specific policy and industry developments, as well as in recognition of the lower penetration of stock exchanges in the economies of the European countries.

In terms of sectors we allow all sectors to our sample except the Nonequity Investment Instruments, which are a separate sector in Datastream, and Financials, which in such tests are usually excluded as a sector that pools together the risks of other sectors.

¹⁰⁴ "NASDAQ" originally stood for National Association of Securities Dealers Automated Quotations

¹⁰⁵ The former name of NYSE MKT was the American Exchange (AMEX)

To recapitulate, the filters employed in Datastream to construct the list of companies used in this study were as follows: “Market” = United States; “Currency” = US dollar; “Exchange” = New York or NASDAQ or NYSE MKT; “Instrument Type” = equity; “Sector” = all sectors except “Nonequity Investment Instruments” and “Financials”.

Available price data start from the beginning of 1973. Values for many of the features are missing, and roughly three quarters of the data is lost due to lack of information on one of the following items: unadjusted price (UP), number of shares (NOSH), price to book value (PTBV). In particular, series for price-to-book value start from January 1980.

Market capitalisation for each security is calculated as the product of unadjusted price (UP) and number of shares issued (NOSH).

A number of securities in Datastream have reported prices but apparently are not traded actively. Their presence in the dataset could affect the estimated return distributions. In order to mitigate the problem of presence of non-traded securities and bearing in mind that the likelihood the price of a traded share to be exactly the same as at the beginning of the period is very close to nil¹⁰⁶, we excluded all monthly returns where the price at the end of the period equals exactly the price at the start of the period unless there is reported positive traded volume of that security (Datastream code 'VO') in that month, i.e. $VO > 0$ (cases where VO is missing were treated as zero volume).

¹⁰⁶ For a random variable with continuous distribution function the probability of some specific value occurring is exactly zero. Stock prices are discrete, but the step is very small so they are close to continuous, hence the probability of exactly the same price occurring at the end of month is rather small (even if the expected return for the month is zero), albeit not zero as in the continuous variable case. However, as long as such incidences of removal of returns are few and can be assumed to occur at random, they should not affect our conclusions, and so this procedure seems reasonable in order to ensure quality of the data set.

Table 4: Average sector weights, Pearson correlations of sectors with the market, and cross-sector correlations

The table shows average weights (“Avg. Weight”), Pearson correlation with overall market (“Corr. with market”) and correlation matrix between the sectors of the stock issuers. The data set includes securities from NYSE, AMEX and NASDAQ. The sub-indices are value-weighted.

	Avg. Weight	Corr. with market	Basic Materials	Consumer Goods	Consumer Services	Financials	Healthcare	Industrials	Oil and Gas	Other and Unclassified	Technology	Telecommunications	Utilities
Basic Materials	0.057	0.824	1	0.723	0.731	0.733	0.572	0.859	0.655	0.61	0.595	0.451	0.447
Consumer Goods	0.079	0.851		1	0.831	0.819	0.782	0.816	0.496	0.631	0.548	0.535	0.594
Consumer Services	0.083	0.892			1	0.822	0.69	0.872	0.424	0.69	0.703	0.592	0.453
Financials	0.057	0.875				1	0.7	0.839	0.522	0.615	0.58	0.56	0.568
Healthcare	0.1	0.788					1	0.715	0.442	0.479	0.588	0.487	0.472
Industrials	0.047	0.95						1	0.612	0.711	0.765	0.562	0.503
Oil and Gas	0.125	0.664							1	0.373	0.42	0.34	0.513
Other and Unclassified	0.006	0.676								1	0.495	0.407	0.414
Technology	0.119	0.81									1	0.498	0.234
Telecommunications	0.252	0.648										1	0.468
Utilities	0.076	0.561											1

Source: author's calculations

3.4.2. Calculated covariates

Excess returns ($ER_{i,t}$) for security i at time t are calculated as the difference between the arithmetic return for the month and the interest on 3-month Treasury bills – middle rate (Datastream code FRTBS3M). The 3-month series is preferred to 1-month T-bills on practical grounds – the latter is available only for dates starting July 31, 2001. The formula for calculation of excess returns therefore is:

$$ER_{i,t} = \frac{P_{i,t+1} - P_{i,t}}{P_{i,t}} - rf_t,$$

where $P_{i,t}$ is the adjusted price at time t for asset i , and rf_t is the risk-free rate from Kenneth French's data library.

Return indices are calculated using value-weights for all securities in our sample for which there is information about market capitalisation. Table 4 lists the Pearson pairwise correlations across constituent sectors, and the average weight of each sectors. Because of the inclusion of NASDAQ in our sample, the weight of telecommunications and technology is high, with a quarter of all companies in the telecommunications business and more than 11% in technology stocks.

As should be expected, the calculated value-weighted market excess return co-varied closely with the excess returns on the commonly used market indices. The correlation with S&P 500 is 0.9900, with Dow Jones Industrials – 0.9318, with NASDAQ Composite – 0.9035, with NYSE Composite – 0.9793, and with MSCI USA – 0.9876. These values are consistent with expectations: the highest correlation was obtained for S&P 500, which is a value-weighted index of 500 leading stocks, and sometimes used as a proxy for the market index as a whole; similarly, the correlation with MSCI USA is also very high, consistent with the fact that it is a free float adjusted market capitalisation index, geared towards large and mid-cap US equities. The correlation with the Dow Jones Industrial Average, on the other hand, is somewhat lower, consistent with the significant differences in coverage and calculation (Dow Jones is a price-weighted index of 30 leading stocks). The correlation with NASDAQ Composite is lower as that index includes both US and non-US stocks (and equity-like instruments like ADRs, REITs, limited partnership interests, etc.) listed at the NASDAQ market.¹⁰⁷ The correlation between the value-weighted excess returns calculated by us and those in *Kenneth R. French Data Library* (2015) is 0.9981.

¹⁰⁷ We discarded NASDAQ non-US equities from our sample.

Most studies in the field employ the CRSP prices augmented with Compustat financial data. The use of a single data set is known to introduce risk of data mining (data snooping). In that respect, the use of an alternative data source adds value to this study. On the other hand, the data coverage of the Datastream database raised some concerns in terms of poorer coverage in the earlier period as well as data errors. Ince and Porter (2006) investigate such differences between Datastream and CRSP/Compustat, identifying the strengths and weaknesses of Datastream data. They also propose filters in order to improve the quality of the estimates. In line with their suggestions, we implement those and some further measures in our sample. With appropriate filtering, Ince and Porter (2006) demonstrate that returns and moments based on Datastream close prices are very similar in magnitude to, and correlate highly with, those based on CRSP data. Nonetheless, we use market excess returns from Kenneth French's online data library in order to limit differences caused by different coverage in the initial sample period. In our study we include only equities; all non-equity instruments like American depository receipts (ADRs), non-equity investment instruments, real estate investment trusts (REITs), shares of beneficial interest, preferred shares, and other non-equity instrument types are excluded. To limit the impact of data errors we have discarded from our sample returns exceeding 300%.

Market capitalisation for each security is calculated as the product of unadjusted price (UP Datastream series) and number of shares issued (NOSH series). $\ln(B/M)$ is the natural logarithm of the ratio of book value of equity (M/B) at the end of the preceding month to the market price of equity¹⁰⁸, and is calculated from the price-to-book value series from Datastream (PTBV) where book values of equity are taken at a lag of six months to ensure that they are known to investors.

Ret_t is the raw month-on-month return for each security at month t , and $XRet_t$ is the excess return for each security, calculated as the difference between Ret and the risk-free rate for the respective month as retrieved from Kenneth French's data library. $IRet_t$ is the idiosyncratic return calculated as the residual corresponding to month t from the four-factor Fama–French–Carhart model estimated with monthly data from $(t - 60)$ to $(t - 1)$.

¹⁰⁸ Practitioners usually use the reciprocal ratio of market capitalization to book value (price-to-book value). Since book value could be very close to zero, we follow to custom in academic literature to place book value in the numerator. We have also excluded cases with negative book value of equity, because for them the logarithm is not defined.

Table 5: Descriptive statistics, 1/1980–3/2013

The table summarises the descriptive statistics for the sample used in the tests. ‘Ret (%)’ is the arithmetic return on the stock; ‘XRet (%)’ is the excess return on the stocks over the risk-free rate for the respective month as retrieved from Kenneth French’s data library; ‘IRet (%)’ is the idiosyncratic return calculated as the residual corresponding to month t from the four-factor Fama–French–Carhart model estimated with monthly data from $(t-60)$ to $(t-1)$.

‘Beta’ is the stock beta calculated by portfolios constructed as in Fama and French (1992). ‘ln(Cap)’ is the natural logarithm of market capitalisation calculated as number of shares times the unadjusted price. ‘ln(B/M)’ is the natural logarithm of book value of equity as available at the end of the preceding month to market price of equity (B/M), and is calculated from the price-to-book value series from Datastream (PTBV) where book values of equity are taken at a lag of six months to ensure that they are known to investors. ‘Ret(-2, -7)’ is the cumulative return for the six months from $(t-7)$ to $(t-2)$; $(t-1)$ is not included in order to control for return reversals. ‘Roll’ is the bid-ask spread calculated Roll’s model; Roll (1984).

‘Mean (EW)’ and ‘Mean (VW)’ are the equally-weighted and the value-weighted values of the respective indicators. ‘St.dev.’ is the standard deviation; ‘Median’, ‘Q1’ and ‘Q3’ are the median and the first and third quartiles of the sample. ‘Skewness’ is the skewness coefficient for the sample. ‘Obs’ is the number of rows for which data is available.

Variables	Mean (EW)	Mean (VW)	St.dev.	Median	Q1	Q3	Skewness	Obs
Ret (%)	1.39	0.58	14.54	0.51	-5.79	7.34	1.75	863,999
XRet (%)	1.05	0.32	14.54	0.19	-6.15	7.01	1.75	863,999
IRet (%)	-0.18	-0.64	12.32	-0.64	-6.29	5.07	1.51	863,993
Beta	1.20	1.00	0.31	1.18	0.98	1.44	0.08	863,999
ln(Cap)	5.87	9.81	1.91	5.76	4.45	7.15	0.31	863,999
ln(B/M)	-0.71	-1.25	0.77	-0.63	-1.12	-0.22	-0.78	863,999
Ret(-2, -7)	1.09	1.09	0.41	1.04	0.88	1.23	3.23	862,210
Roll	6.88	4.28	4.24	5.81	4.08	8.4	2.74	863,970

Source: author’s calculations

The CAPM proposes that the only factor explaining the cross-section of market returns is the market beta. The estimation of beta, however, is fraught with problems, and a principal one among those is the measurement errors in betas. In this study we measure systemic risk (*Beta*) following the procedure proposed by Fama and French (1992) with minor modifications. In each month we calculate the betas using the previous 60 months (but no less than 30 months) of data, not including the current month. Stocks are then assigned to five quintile-size portfolios. Within each quintile-size portfolio, ten decile beta portfolios are formed. For calculation of the quintile breakpoints for size sorting and the decile breakpoints for beta sorting we exclude the NASDAQ and NYSE MKT shares (i.e., NYSE breakpoints). The beta for each of the 50 size-beta portfolios is calculated running full-period regression of the equally-weighted average monthly excess returns on the current and previous period market excess returns. Then *Beta* was calculated as the sum of the slopes of the two market returns, which is intended to correct for non-synchronous trading, and that beta is assigned to all securities in those portfolios.

Jegadeesh and Titman (1993) report that past returns are predictors of current performance (momentum effect), and following Fu (2009) we construct a return proxy equal to the cumulative gross return from months $t - 7$ till $t - 2$ inclusive, i.e. $Ret(-2, -7) = \prod_{i=2}^7 (1 + R_{t-i})$, where R is the gross return; thus a cumulative decline of 20 per cent over the six-month period is recorded as $Ret(-2, -7) = 0.8$. Return in $t - 1$ is not included, in order to ensure that the variable significance is not due to return reversals. Instead, return in month $t - 1$ is used as a separate independent variable.

Asset liquidity is discussed in the context of the cross-section of stock returns. Spiegel and Wang (2005) demonstrate that liquidity is inversely correlated with idiosyncratic risk. There is no single measure of volatility, and a number of measures are used in the literature. For example, Amihud and Mendelson (1986) propose that liquidity measured in terms of the bid-ask spread is desired by investors and therefore the less liquid shares should earn a premium. Fu (2009) measures idiosyncratic risk in terms of rolling average trading volume, as well as the coefficient of variation of traded volume. We estimate liquidity using Roll (1984) model-based estimator (*Roll*) of the spread, because the spread between the buy and sell sides of the market might be a less ambiguous measure of liquidity compared to traded volume, where arguably the free float instead of the total number of shares should be in the denominator. Roll (1984) proposes that the spread (S) could be estimated from the

auto-covariance of prices as they bounced back and forth between the sell side and the buy side of the market:

$$S = \begin{cases} 2\sqrt{-Cov(\Delta P_t, \Delta P_{t-1})} & \text{if } Cov(\Delta P_t, \Delta P_{t-1}) < 0, \\ 0 & \text{otherwise} \end{cases},$$

where $Cov(\cdot)$ is the first order return covariance. The estimator is biased when the sample size is small and the frequency is low (less frequent than daily); therefore we estimate the spread using the daily returns from the past year.

The descriptive statistics reported in Table 4 overall are similar to those reported in other studies like Fama and French (1992) and Fu (2009), which confirms that our sample is representative for the market.

Earlier studies such as that of Fama and MacBeth (1973) identify idiosyncratic risk as the standard deviation of the residuals from the fitted market models:

$$\sigma_{ols,t}^2 = \sum_{k=1}^T \frac{(\varepsilon_{i,t-k})^2}{T - df},$$

where $\varepsilon_{i,t-k}$ denotes the idiosyncratic residual from the FFC model for stock i in month $(t - k)$ obtained from a regression estimated over the period $(t - 60)$ until $(t - 1)$. T is the actual number of months for which information is available; thus T is between 60 and 30 (the minimum number of months required for estimation); df was the residual degrees of freedom for the estimated model and equals the number of estimated parameters, so that $df = 5$ (four factors and an intercept). It is immediately clear from the specification that the return for period t plays no role in the estimation of OLS idiosyncratic variance, $\sigma_{ols,t}^2$. Clearly, the resulting estimates are strongly autocorrelated as the estimates are obtained using nearly-identical data sets – in two consecutive months there would be one month leaving the sample and one month entering the sample (rolling window design), or one period added to the sample (expanding window design). In our study idiosyncratic volatility (standard deviation) is obtained from the residuals of the rolling regressions with data ending at period $(t - 1)$ and including 60 to 30 months of return (as available). These volatility estimates are denoted by IV_{t-1}^{OLS} ; to avoid unnecessarily cluttering the notation we sometimes omit the time subscript, which is set to $(t - 1)$ to remind us that the return at period t was not incorporated in that estimate.

The second measure of expected idiosyncratic volatility is estimated using GARCH(1,1) model with SGED innovations. The forecasts from that model are denoted by \widehat{IV}_t^{garch} . The GARCH model is fitted using an expanding window containing all available

data up to the beginning of the month (i.e. end of $t - 1$); similarly to Fu (2009), at least 60 months of data¹⁰⁹ were required in order to estimate the model and yield a forecast for month- t idiosyncratic volatility. No information from month- t is used in the estimation of any parameters of the model. Variance forecasts are obtained from the model fitted until $(t - 1)$. Essentially this means that if idiosyncratic volatility is above its mean-reverting level, the model *forecasts* for month t would be lower compared to the *fitted* variance for month $(t - 1)$; the opposite would hold if month- $(t - 1)$ variance is below the mean-reverting level. The existence of the mean-reverting level is ensured by the parameter constraints on the GARCH model ($\alpha + \beta < 1$, $\alpha > 0$, $\beta > 0$). Thus our model is specified as follows:

$$\begin{aligned} (r_{i,t} - r_{f,t}) &= \beta_0 + \beta_1(r_{M,t} - r_{f,t}) + \beta_2r_{SMB,t} + \beta_3r_{HML,t} + \beta_4r_{MOM,t} + \varepsilon_{i,t}, \\ \varepsilon_{i,t} &= \sqrt{\sigma_{i,t}^2}v_{i,t}, \\ v_{i,t} &\sim SGED, \mathbb{E}v_{i,t} = 0, \mathbb{E}v_{i,t}^2 = 1 \\ \sigma_{i,t}^2 &= \omega + \alpha\varepsilon_{i,t-1}^2 + \beta\sigma_{i,t-1}^2, \\ \alpha &> 0, \beta > 0, \alpha + \beta < 1. \end{aligned}$$

The estimation of the parameters of the above model proceeds sequentially. First, the mean equation is estimated using OLS over the period from month 1 to month $(t - 1)$, if at least 60 months of continuous data were available (expanding window design). The estimated series of residuals ($\varepsilon_{i,t}$, $t = 1 \dots (t - 1)$) is then used to estimate the maximum-likelihood parameters of the volatility model. The fitted model produces the forecasts for month t using the last idiosyncratic innovation for month $(t - 1)$ and the fitted variance for month $(t - 1)$.

¹⁰⁹ In cases where a stock was not traded in a particular month, identified by either trading volume equal to zero or by zero change of the price in the month. Such deletions resulted in gaps of the series, and thus some infrequently traded securities were excluded from the criterion of 60 months of contiguous trading. This alleviates concerns that the finding of significance of idiosyncratic risk was driven by few securities with thin trading and are not representative for the market.

Table 6: Descriptive statistics for idiosyncratic volatility forecasts, 1/1980–3/2013

The table reports the descriptive statistics for the calculated volatility forecasts. ‘ IV^{OLS} ’ is the volatility of the residual of rolling monthly regressions using thirty to sixty months of data, as available. ‘ IV_{t-1}^{daily} ’ is the idiosyncratic volatility from the daily data in the preceding month calculated using the residuals from Fama-French regressions with the daily data for the preceding month; for convenience the estimates are scaled to monthly frequency using \sqrt{T} . ‘ \widehat{IV}_t^{garch} ’ is the forecasts from GARCH(1,1) model with SGED innovations estimated using expanding window comprising of at least sixty months of continuous data. ‘ \widehat{IV}^{arma} ’ is expected volatility from ARMA(1,1) model fitted on the available series of IV_{t-1}^{daily} . ‘ m ’ is the mean-reverting level of volatility implied by the fitted ARMA model.

‘Mean (EW)’ and ‘Mean (VW)’ are the equally-weighted and the value-weighted values of the respective indicators. ‘St.dev.’ is the standard deviation; ‘Median’, ‘Q1’ and ‘Q3’ are the median and the first and third quartiles of the sample. ‘Skewness’ is the skewness coefficient for the sample. ‘Obs’ is the number of rows for which estimates are available.

Variables	Mean (EW)	Mean (VW)	St.dev.	Median	Q1	Q3	Skewness	Obs
IV^{OLS} (%)	12.22	7.64	6.72	10.54	7.51	15.19	1.63	863,993
$\ln(IV^{OLS})$	2.37	1.94	0.50	2.36	2.02	2.72	0.18	863,993
IV_{t-1}^{daily} (%)	10.98	6.41	7.81	8.82	5.84	13.63	2.39	729,628
$\ln(IV_{t-1}^{daily})$	2.20	1.70	0.62	2.18	1.76	2.61	0.15	729,628
\widehat{IV}_t^{garch} (%)	11.39	7.56	5.67	10.08	7.34	14.08	1.57	746,953
$\ln(\widehat{IV}_t^{garch})$	2.32	1.94	0.46	2.31	1.99	2.64	0.16	746,953
\widehat{IV}^{arma} (%)	11.61	6.58	6.45	9.97	7.00	14.52	1.64	812,920
$\ln(\widehat{IV}^{arma})$	2.32	1.79	0.51	2.30	1.95	2.68	0.17	812,920
m (%)	13.41	7.68	6.93	11.84	8.25	16.90	1.43	812,793
$\ln(m)$	2.48	1.95	0.49	2.47	2.11	2.83	0.13	812,793
Spread	-0.20	-0.21	0.49	-0.24	-0.56	0.02	1.12	812,732

Source: author’s calculations

IV_{t-1}^{daily} is the previous month idiosyncratic volatility, calculated as the standard deviation of daily idiosyncratic residuals from the three-factor Fama–French model¹¹⁰ fitted on daily data in the previous calendar month ($t - 1$). We require at least 15 non-zero returns in the calculation month, save for September 2001, where only 12 trading days are required. Zero-return days are discarded in order to reduce the impact of infrequent trading on the volatility measure. IV_{t-1}^{daily} is scaled to monthly frequency by multiplying by square root of the number of trading days in the respective month in order to ease comparison of regression slopes across volatility measures.

The last measure of expected idiosyncratic volatility are the forecasts from the history of IV_{t-1}^{daily} produced using ARMA(1,1) model:

$$h_t - m = \phi(h_{t-1} - m) + \theta\varepsilon_{t-1} + \varepsilon_t,$$

where m , ϕ and θ are the parameters of mode. Parameter m is also called the mean-reverting level of volatility because when $\phi \in (0,1)$, the forecasted values of h would converge to m , given enough time. In general, the values of $\phi, \theta \in (-1,1)$ ensure that the ARMA process is invertible.¹¹¹ The unconditional expected value of the process equals m , $\mathbb{E}(h_t) = m$, and ϕ controls the speed of mean-reversion. An alternative parametrisation for the process uses $\mu = m(1 - \phi)$ instead of m , giving the alternative form $h_t = \mu + \phi h_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t$.

As particular cases the specification admits constant expected volatility ($\phi = 0, \theta = 0$) and random-walk volatility ($\phi = 1, \theta = 0$). Stationarity requires that the parameters ϕ, θ of ARMA(1,1) be in the interval $(-1, +1)$. However, a negative value of ϕ (the mean-reversion parameter) implies that expected volatility oscillates around the mean-reverting level. Therefore we enforce the stricter limit for the mean-reversion parameter: $\phi \in [0,1)$. A downside of the ARMA(1,1) approach is that the forecasted volatility may become negative. We discard from our sample those forecasts where volatility is non-positive (109 records) or exceeds 200% on a monthly basis (8 records), which is 0.01% of the sample total of 812,920 forecasts.

¹¹⁰ The preference to the three-factor model was based on two considerations. Firstly, to be consistent with the approach of Ang et al. Secondly, in each month we have at most 23 daily returns, which warrants more parsimonious specification.

¹¹¹ The formulation of the ARMA process means that all observed values of h_t are function of the history of unobserved errors. If the process is invertible, then these errors can be represented as weighted sums of the observed realisations.

Overall, the measures of idiosyncratic risk that we considered in this study are fairly representative of the prevailing practice. Thus IV^{OLS} represents the filter-based measures; other examples of this class could be the Hodrick-Prescott filter used by Cao (2010) and Cao and Xu (2010), or $IV^{monthly}$ – the moving average of IV^{daily} , employed by Bali and Cakici (2008). GARCH(1,1) could be construed as representative of the GARCH-based models. IV_{t-1}^{daily} , the measure used by Ang et al. (2009, 2006) is a measure based on assumption of random-walk volatility, and in that respect also related to the Integrated GARCH model. Finally, \widehat{IV}^{arma} is representative of the mean-reversion volatility hypothesis (stationary GARCH also falls in this category). We consider these four measures as forecasts for period- t idiosyncratic variance. Table 5 provides summary statistics for the idiosyncratic volatility measures.

The covariates employed in the studies of the cross-section of returns tend to be correlated. Such correlation is not only empirical regularity, but is also predicted from various theories. In Table 6 we provide summary statistics for the key idiosyncratic risk covariates broken down by the ten decile beta portfolios and the five quintile capitalisation portfolios. The table confirms that stocks with higher capitalisation have lower betas and lower idiosyncratic risk, and indeed the mean idiosyncratic risk of the high-capitalisation stocks was roughly half that of the low-capitalisation stocks.

Table 7: Mean volatilities by portfolios sorted by beta and capitalisation, 1980/7–2013/3

The table reports the average volatilities of ten decile portfolios sorted by market beta (Panel A) and five quintile portfolios formed by market capitalisation (Panel B). ‘ IV^{OLS} ’ is the volatility of the residual of rolling monthly regressions using 30 to 60 months of data, as available. ‘ IV_{t-1}^{daily} ’ is the idiosyncratic volatility from the daily data in the preceding month calculated using the residuals from Fama-French regressions with the daily data for the preceding month. ‘ \widehat{IV}_t^{garch} ’ is the forecasts from GARCH(1,1) model with SGED innovations estimated using expanding window comprising of at least sixty months of continuous data. ‘ \widehat{IV}_t^{arma} ’ is expected volatility from ARMA(1,1) model fitted on the available series of IV_{t-1}^{daily} .

The table demonstrates the significant correlation between market capitalisation, beta and the various volatility forecasts. Thus, the volatility of the stocks in the highest-beta decile portfolio is between 5.45 and 7.77 percentage points higher compared to the lowest-beta portfolio. Similarly, the volatility of the lowest-capitalisation quintile portfolio is between 7.08 and 9.08 percentage points higher than the volatility of the highest-capitalisation stocks.

Portfolio	$XRet$	$Beta$	$\ln(Cap)$	IV^{OLS}	\widehat{IV}_t^{garch}	IV_{t-1}^{Ang}	\widehat{IV}_t^{arma}
<i>Panel A: Portfolios sorted by Beta</i>							
Low-beta	0.69	0.77	12.72	10.42	9.60	9.52	9.98
2	0.86	0.82	12.78	9.46	9.14	9.21	9.65
3	0.94	0.97	12.88	9.51	9.18	9.45	9.75
4	0.95	1.03	12.89	9.94	9.48	9.80	10.12
5	0.92	1.10	12.81	10.59	10.10	10.38	10.74
6	1.14	1.17	12.83	11.11	10.15	10.77	10.88
7	1.18	1.26	12.83	11.72	10.71	11.32	11.36
8	1.02	1.33	12.82	12.43	11.30	11.85	11.92
9	1.20	1.44	12.77	13.83	12.06	13.01	12.80
High-beta	1.36	1.70	12.61	18.19	14.42	16.00	15.43

Portfolio	$XRet$	$Beta$	$\ln(Cap)$	IV^{OLS}	\widehat{IV}_t^{garch}	IV_{t-1}^{Ang}	\widehat{IV}_t^{arma}
<i>Panel B: Portfolios sorted by capitalisation</i>							
Low-cap	1.46	1.29	11.19	15.60	14.81	14.46	15.74
2	0.92	1.22	12.68	11.82	10.40	11.14	10.78
3	0.79	1.18	13.46	10.19	9.00	9.78	9.21
4	0.71	1.11	14.36	8.85	7.91	8.71	8.03
High-cap	0.52	0.99	15.90	7.27	6.62	7.38	6.66

Source: author's calculations

3.5. Classification of volatility regimes

The last methodological aspect that we would like to explore in this chapter is the identification of high-volatility episodes. The purpose of such classification of time periods is to examine the robustness of our results. For example, some of the relationships could hold only in low-volatility state, or the significance of some results could be due to outliers occurring in a high-volatility environment.

The approach we take to identify such episodes is based on Fink et al. (2010), who identify high-volatility episodes in their sample by using a Markov chain with two states, where volatility in each state is assumed to be normally distributed with unknown mean and variance.

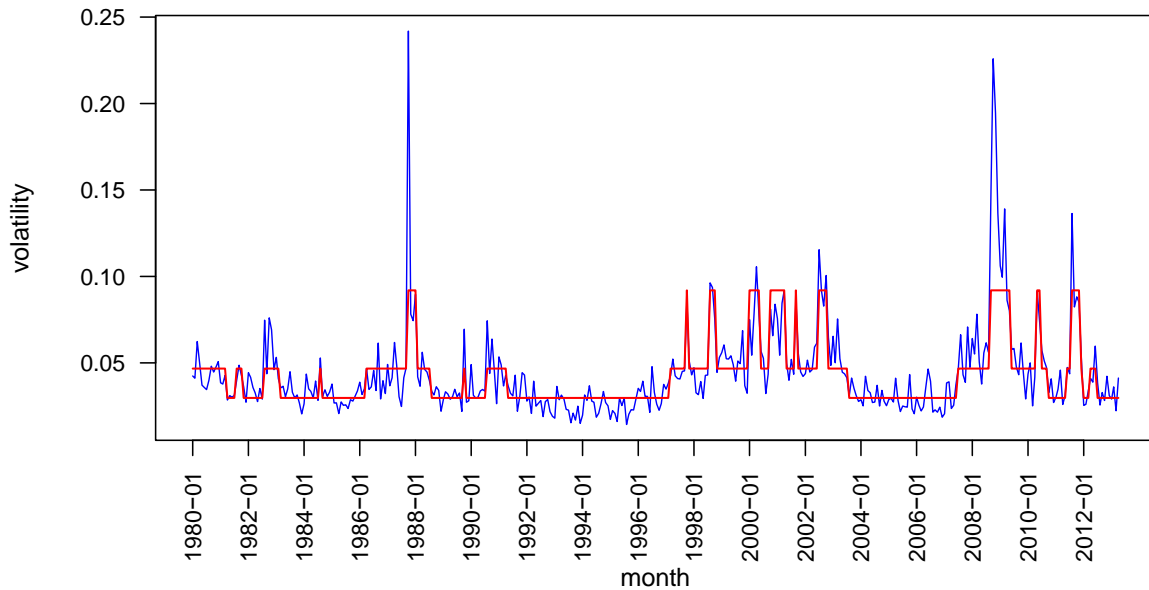
Formally, let Π be a right stochastic¹¹² matrix of transition probabilities whose element $\pi_{i,j}$ is the probability of transition from state i to state j conditional on the system being in state i . The list of states is complete, i.e. there is no other state in which the system might be, hence the rows of the transition matrix should sum to 1. They assume that when the market volatility is in state i , market volatility is independently and identically normally distributed with some unknown parameters (that are to be estimated), i.e. $\sigma_{market}N(\mu_i, \sigma_i)$. The parameters of that system (the matrix Π and the pairs $(\mu_i, \sigma_i), \forall i$) could be estimated using the Baum–Welch algorithm (Baum et al., 1970), and the sequence of states of the market (the Viterbi path) at any point in time could be estimated using the Viterbi (1967) algorithm.

Firstly, we estimate the above model with 2 states.¹¹³ We calculate monthly volatilities of the DataStream Return Index (all sectors) from daily data as $\sigma_{m,t}^2 = \frac{21}{N} \sum_{i \in t} R_i^2$, where i ran over the days in month t , and N was the number of days in month t . The calculated volatility of the index is displayed in Figure 2; the mean volatility is 0.04267, while the first, second and the third quartiles respectively equal 0.02868, 0.03637, and 0.04809. The maximum volatility of 0.2419 is observed in October 1987, and the second-highest volatility of 0.2259 occurred in October 2008.

¹¹² A square matrix is right stochastic matrix if all its elements are non-negative real numbers, with each row summing to 1

¹¹³ We implemented these calculations using the ‘RHmm’ package of R (Taramasco and Bauer (2013))

Figure 3: Realised volatility of the market and volatility of the Hidden Markov Model with Three States



Source: author's calculations

The estimates from the Hidden Markov Model with 2 states and normal distribution of volatilities are given in Table 8. The results suggest that the mean volatility in the high-volatility state (state 1) is more than twice the market volatility in the low-volatility state (state 2). Furthermore, the dispersion in the high-volatility state is in order of magnitudes higher than in the low volatility state. This result is consistent with the observed high peaks in the volatility chart, which imply that either higher number of states might be appropriate, or conditional distributions with heavier tails might be justified.

We further note that out of 484 months used in the estimation (from January 1973 until April 2013), the market is in the high-volatility state for 125 months, which constitutes more than a quarter of that period. Therefore, to further narrow down the list of highly volatile episodes, we estimate the same model with three states. The results are presented in Table 8; the high-volatility state prevails in 41 of the months, and the medium-volatility state – in another 181 months. The mean volatility in the high-volatility state is more than thrice the mean volatility in the low-volatility state and nearly twice as high as the volatility in the medium state. Interestingly, the transition matrix suggests that one should not expect transitions directly from state 1 to state 3, or from state 3 to state 1; such transitions can occur over the course of at least two months.

Table 8: Parameter estimates for Hidden Markov Model with two states and Normal distribution of volatilities in each state

The table reports the parameters of a Hidden Markov Model fitted on the series of total market returns. The models assumed the existence of two volatility regimes – high-volatility (regime 1) and low-volatility (regime 2). In each regime volatility is assumed to be independently and identically normally distributed with mean and variance of μ_i and σ_i^2 . The transition matrix is $\Pi = [\pi_{i,j}]$, where $\pi_{i,j}$ is the probability of transition from state i to state j .

‘Estimate’ is the point estimate of the respective parameter, while ‘Std. Error’, ‘t value’ and ‘ $Pr(> |t|)$ ’ are the standard error, the t-statistic and the p-value of the estimate. Parameters are estimated using the Baum–Welch algorithm.

The result shows the existence of at least two clearly separated volatility regimes, as evidenced by the significant difference between μ_1 and μ_2 compared to the standard errors of the two estimates, and with significant volatility persistence as measured by $\pi_{i,i}$, which range between 89.71% and 92.26%.

	Estimate	Std. Error	t value	$Pr(> t)$
$\pi_{1,1}$	0.8971	0.0345	26.012	0.0000
$\pi_{1,2}$	0.1029	0.0345	2.985	0.0028
$\pi_{2,1}$	0.0374	0.0119	3.146	0.0017
$\pi_{2,2}$	0.9626	0.0119	81.059	0.0000
μ_1	0.0686	0.0033	20.514	0.0000
σ_1^2	0.0011	0.0000	238.109	0.0000
μ_2	0.0333	0.0006	59.535	0.0000
σ_2^2	0.0001	0.0000	92.141	0.0000
Log Likelihood:	1363.13			
BIC Criterion:	-2682.98			
AIC Criterion:	-2712.26			

Source: author's calculations

Table 9: Parameter estimates for Hidden Markov Model with three states and Normal distribution of volatilities in each state

The table reports the parameters of a Hidden Markov Model fitted on the series of total market returns. The models assumed the existence of three volatility regimes – high-volatility (regime 1) and low-volatility (regime 3). In each regime volatility is assumed to be independently and identically normally distributed with mean and variance of μ_i and σ_i^2 . The transition matrix is $\Pi = [\pi_{i,j}]$, where $\pi_{i,j}$ is the probability of transition from state i to state j .

‘Estimate’ is the point estimate of the respective parameter, while ‘Std. Error’, ‘t value’ and ‘ $Pr(> |t|)$ ’ are the standard error, the t-statistic and the p-value of the estimate. Parameters are estimated using the Baum–Welch algorithm.

The result shows the existence of three clearly separated volatility regimes, as evidenced by the significant difference between μ_1 , μ_2 and μ_3 compared to the standard errors of the estimates, and with significant volatility persistence as measured by $\pi_{i,i}$, which range between 72.80% for the high-volatility regime and 93.03% in the low-volatility regime.

	Estimate	Std. Error	t value	$Pr(> t)$
$\pi_{1,1}$	0.7280	0.1185	6.144	0.0000
$\pi_{1,2}$	0.2720	0.0897	3.032	0.0024
$\pi_{1,3}$	0.0000	0.1495	0.000	1.0000
$\pi_{2,1}$	0.0798	0.0244	3.270	0.0011
$\pi_{2,2}$	0.8228	0.0364	22.602	0.0000
$\pi_{2,3}$	0.0975	0.0279	3.498	0.0005
$\pi_{3,1}$	0.0000	0.0150	0.000	0.9998
$\pi_{3,2}$	0.0697	0.0193	3.611	0.0003
$\pi_{3,3}$	0.9303	0.0199	46.659	0.0000
μ_1	0.0919	0.0074	12.404	0.0000
σ_1^2	0.0015	0.0000	86.926	0.0000
μ_2	0.0466	0.0008	59.801	0.0000
σ_2^2	0.0001	0.0000	71.917	0.0000
μ_3	0.0297	0.0005	56.319	0.0000
σ_3^2	0.0000	0.0000	70.808	0.0000
Log- Likelihood:	1440.51			
BIC Criterion:	-2794.47			
AIC Criterion:	-2853.02			

Source: author's calculations

Table 10: Periods with high market volatility

The table summarises the episodes of high market volatility as identified by the three-state hidden Markov chain model. The model is estimated with the Baum–Welch algorithm, and the sequence of states is estimated using the Viterbi algorithm. Months with high volatility have estimated mean volatility of 9.19% with standard deviation of 3.87%. Brief comments on the market events that were associated with the market turbulence are added for clarity.

Start	End	Comments
July 1974	October 1974	The crisis of 1973-1974 followed shortly after the termination of the convertibility of US dollars to gold, one of the tenets of the Bretton Woods system, and the start of the first oil crisis in October 1973. The stock market crash started in November 1973 with DJIA losing 14% in that month and S&P 500 losing 7%; the fall lasted 12 month ending October 1974, with DJIA and S&P 500 losing 30.4% and 36.8%. ¹¹⁴
October 1987	January 1988	On October 19, 1987, the Dow Jones recorded the largest fall in a single day, falling 22.6%. S&P 500 fell 12.1% in October and 12.5% in November. The fall resulted in brokers extending significant credits to their customers to finance margin calls. The Fed stepped in to support the financial system by lending to banks, expanding the monetary base. Throughout the remainder of the year stock prices remained volatile. ¹¹⁵
October 1997	October 1997	A short volatility episode related to the Asian financial crisis. Consequently, on October 27, 1997, DJIA fell 7.18%, scoring its twelfth biggest daily percentage loss. The shock was short-lived, however, and on the next day DJIA recovered more than 60% of the previous day loss.
August 1998	October 1998	On August 17, 1998 the government of the Russian Federation devalued the Ruble, defaulted on domestic debt, and declared a 90-day moratorium on external debt repayment. At the end of August, DJIA fell 11.5% in three days, and the market remained depressed until being stimulated by a series of interest rate cuts in October

¹¹⁴ Mishkin and White (2002)

¹¹⁵ Ibid.

Start	End	Comments
January 2000	May 2000	The dot-com bubble of 1997–2000 reached its peak on March 10, when NASDAQ Composite peaked at more than twice its value a year earlier. The rapid decrease in value of dot-com shares continued until 2001 as many tech companies that had spent their cash but were failing to make a profit found it difficult to raise more funds on the stock exchange.
October 2000	April 2001	The period was marked by continued decline of tech stocks coupled with a slowing economy that moved into recession in March 2001
September 2001	September 2001	A short-term volatility following the terrorist attacks on September 1, 2001
July 2002	October 2002	The episode was a continuation of the bear market that started in 2000, with both NASDAQ and DJ scoring three consecutive years of losses. In July 2002 volatility surged when DJ fell for 11 of 12 consecutive days, and then slid to 7700, then recovered to just above 9000 in August, and then fell back to below 7300 in October.
September 2008	May 2009	In September 2008 the subprime financial crisis intensified; in early September the US federal government had to bail out Fannie Mae and Freddie Mac, guarantors for many sub-prime mortgages. On September 15, Lehman Brothers, an investment bank with large exposure to subprime assets, filed for bankruptcy, prompting worldwide financial panic. By the end of the month, two more American banks collapsed – Washington Mutual and Wachovia. ¹¹⁶
May 2010	June 2010	Continuing problems of Greece prompted its downgrade by Moody's (to 'A3') and Standard & Poor's (to 'BB-'); at the start of May the Eurozone agreed on a large bailout package for Greece in return for structural reforms and austerity measures

¹¹⁶ Kingsley (2012)

Start	End	Comments
August 2011	November 2011	Continuing problems in the Eurozone (bailout of Portugal in May, second bailout of Greece in July, concerns over risks for contagion to Spain and Italy) and the US (downgrade of the sovereign rating)

Source: the author

The transition probabilities reported in Table 9 provide a good illustration of two of the stylised facts on volatility forecasting formulated by Engle and Patton (2001). For example, the largest entries in the transition matrix are located along the main diagonal ($\pi_{1,1}$, $\pi_{2,2}$, $\pi_{3,3}$). Therefore, the current state is likely to persist into the following period. At the same time, the probability of extreme transitions (from high volatility to low volatility, $\pi_{1,3}$, or from low volatility to high volatility regime, $\pi_{3,1}$) is negligibly low. This structure of the transition matrix provides a demonstration of the persistence property of volatilities. Similarly, volatilities are said to be mean-reverting, and that can be seen in the transition matrix as well. Indeed, if we identify the medium state as the mean level of volatility (and it is the closest of the three states to the unconditional mean), then we see that the second highest transition probabilities of the high- and low- states are those for transition towards the medium state ($\pi_{1,2}$, $\pi_{3,2}$). Another common observation concerning typical patterns of volatility dynamics is also evident in the transition matrix: the periods of low volatility are more persistent than higher volatility states, which in our case results in $\pi_{3,3} > \pi_{2,2} > \pi_{1,1}$. Such a finding is consistent with the observations of Friedman and Liabson (1989), who argue that price movements comprise of ordinary and extraordinary movements, and that their evidence concerning volatility persistence suggests that it is due to the ordinary component of returns.

Table 10 lists the high-volatility periods identified by the hidden Markov model with three states, together with short comments on market events that unfolded in the respective periods and could explain the triggers for the respective turbulence episode.

3.6. Conclusions

In this chapter we motivated our choice of methodology and we described the methodology for performing the tests implemented for this study. The research problem could be tackled by various approaches using a wide range of available methods. We motivated our

choice of a deductive, quantitative, statistical methodology. The choice of a deductive method was motivated by the significant risk of data mining is a larger set of companies were involved in the study, as well as concerns of the external validity of our results in case of a deeper focus on a small set of companies. The focus on external validity, combined with the continuous nature of the explained variable (stock returns), motivated also our choice of statistical methods; the limitation of that approach was the loss of detailed information of the diverse aspects of risk and the risk assessment workflow of institutional and individual investors. On the other hand, that allowed our study to cover a substantial part of the investment universe and to span a period of over three decades.

As a primary data source we use Thomson Reuters Datastream, augmented with factor returns from Kenneth French's database. Idiosyncratic returns are measured relative to the four-factor Fama–French–Carhart model estimated on a rolling window of length 60 months (but not less than 30 months), ending at month $(t - 1)$ in order to avoid look-ahead bias. Idiosyncratic volatilities are calculated in four different methods, based on the prevailing practice, with certain refinements that were motivated in this chapter. In particular, we use the following volatility estimators: (i) idiosyncratic return from monthly OLS regressions; (ii) GARCH(1,1) with Skew Generalised Error Distribution estimated on an expanding window of length at least 60 months and ending at month $(t - 1)$; (iii) daily idiosyncratic volatility from the previous month (Ang et al); (iv) forward-looking estimates of future daily volatility using the past history of estimator (iii).

The predictive accuracy of idiosyncratic volatility measures is evaluated primarily through Mincer-Zarnowitz regressions to avoid scaling between different time frequencies impacting the ranking of predictive accuracy. As true volatilities we use the squared monthly returns and the sum of daily squared returns filtered from a daily EGARCH model fitted on the entire available series for each security.

The Fama–MacBeth methodology is selected to test whether idiosyncratic volatilities predict the cross-section of returns. We use individual securities as assets in the Fama and MacBeth (1973) regressions. Standard errors are calculated using the Newey and West (1987) adjustment with four lags.

Overall, in this chapter we set out our methodology to quantify idiosyncratic risk and to measure its correlation with market returns. The next chapter presents the results from the tests described in this chapter. In particular, it will compare alternative volatility forecasts in terms of their predictive accuracy, the correlation between idiosyncratic risk and stock returns, and robustness tests.

4. Idiosyncratic Risk and the Cross-Section of Stock Returns: Empirical Findings

4.1. Introduction

In this chapter we present our empirical finding concerning the link between idiosyncratic risk and stock returns using the methods presented in Chapter 3. Firstly, we shall examine which of the four idiosyncratic risk estimators from Section 3.4 is the best predictor of future (next-period) idiosyncratic volatility. In the last chapter we suggested that the estimators using measurements from daily returns should offer superior performance, a hypothesis that we test in Section 4.2. Additionally, we shall also test the stationarity of volatility processes. This was previously tested by Fu (2009); our goal in this section is to confirm his findings, and more importantly, to compare qualitatively the forecasting methods. If volatilities are overwhelmingly stationary, we should expect that good future forecasts should share that property. If they do not, then it would seem that the forecasts do not follow faithfully changes in volatilities over time.

In Section 4.3 we shall examine how the various volatility estimators perform as predictors of the cross-section of returns. If expected idiosyncratic risk truly predicts returns, then we should expect that the better predictors of future volatility should result in more reliable forecasting of returns. On the other hand, the empirical tests need to operationalize the concept of idiosyncratic risk from the model economy to the actual financial markets. In the model economy idiosyncratic volatility is fixed and known to all participants. In the financial markets, it is neither constant, nor known. We abstract from the problem how investors learn the volatilities of securities and what the impact of differences in forecasts (non-homogeneous beliefs) is on equilibrium returns. Instead, we focus on how the stationarity of volatilities and their reversion to their means might be incorporated in the investor decision-making and, hence, in returns. In Section 4.3 we shall test if expected volatility for the month predicts the cross-section. From a certain perspective, those tests could be based on the counterfactual assumption that volatilities have unit root and our best forecast of all future volatilities is the next-period volatility, i.e. these tests do not incorporate any information on future volatility other than the one-step forecast.

In Section 4.4 we build on the conflicting evidence from the different models (reported in Section 4.3) and explore how the mean-reverting level of volatility explains the cross-section of returns. In the simple autoregressive context the expected future path of volatilities could be summarised in terms of just three parameters: the starting value of volatility, the mean-reverting level, and the speed of mean-reversion. In Section 4.3 we examine how next-period volatility (which is essentially the starting point of the expected volatility path) predicts the cross-section of returns, and in Section 4.4 we examine the significance of the mean level. It might be supposed that the speed of mean-reversion could also play some part in explaining the cross-section, but our exploratory analysis (not reported here) did not support that hypothesis. Therefore, in Section 4.4 we report how the mean-reverting level of volatility explains the cross-section of returns. In the section we also report various robustness tests for our result. Indeed, an argument could be made that the predictive performance of the mean-reverting level is due solely to some particular subsection of the population, which experiences some short-lived positive or negative momentum. Therefore, we examine whether our results are robust across various subsections of the population, thereby eliminating some hypothesised explanations of our results. Section 4.5 extends those tests further by performing additional tests for omitted factor, using portfolio alphas, as well as daily data. These tests are somewhat different from the other tests reported in this study in terms of methodology. Therefore, they do not rely solely on the methodology laid out in Chapter 3, but also employ additional models relevant only for those sections (e.g. statistical factor analysis and Component GARCH); the specific information on these tests is developed in the section where the test is performed.

Finally, Section 4.6 concludes the presentation of empirical findings.

4.2. Comparison of volatility forecasts

The first question that we address is the quality of forecasts yielded by the selected methods for volatility forecasting in the domain of idiosyncratic risk. Such an analysis requires two inputs: estimates of the *ex ante* expected returns, and measurements of the *ex post* realised volatility.

The rationale and the methodology for calculation of the four selected volatility forecasts is described in the preceding chapter. Let IV^{ols} denote the standard deviation of the residuals from the Fama–French–Carhart model fitted on a rolling window of 60 (but no less

than 30) months ending at $(t - 1)$. Let \widehat{IV}_t^{garch} denote the expected volatility estimated using GARCH(1,1) with SGED (skewed generalised error distribution), where parameters are estimated on expanding windows containing all available data up to the beginning of the month; at least 60 months of data are required. Let IV_{t-1}^{daily} denote the previous month idiosyncratic volatility, calculated as the standard deviation of daily returns in the previous calendar month, requiring at least 15 non-zero returns in the preceding month.¹¹⁷ Let \widehat{IV}_t^{arma} denote the forecast from the ARMA(1,1) model calculated using the available history of IV_{t-1}^{daily} ; the corresponding mean-reverting level implied by the ARMA model is denoted by m . The summary statistics for IV^{OLS} , \widehat{IV}_t^{garch} , IV_{t-1}^{daily} , \widehat{IV}_t^{arma} and m , together with their less-skewed log-transforms, are provided in Table 6 on p. 117.

As an initial exploratory analysis we calculate the cross-sectional correlation between the idiosyncratic volatility forecasts and the proxies for the realised (*ex post*) variance. As proxies for the unobservable true variance we use: (i) the squared idiosyncratic return ($IRet_{i,t}^2$); (ii) the in-sample volatility ($IV_{i,t}^{garch}$) filtered from the full-sample GARCH(1,1) model with SGED shocks, estimated with monthly data; and (3) the mean idiosyncratic volatility filtered in-sample from daily data ($IV_{i,t}^{egarch}$) and averaged by months.¹¹⁸ For each month we calculate the Spearman's rank-correlation coefficient between each pair of proxies of true (*ex post*) volatility and expected (*ex ante*) volatility.¹¹⁹ The preference to the Spearman's rank-correlation over the standard Pearson's coefficient¹²⁰ is driven by a desire to mitigate the possible impact of outliers on the estimated correlation coefficients.

¹¹⁷ IV_{t-1}^{daily} is scaled to monthly frequency by multiplying by square root of the number of trading days in the respective month in order to ease the comparison of coefficients across volatility measures.

¹¹⁸ Expected (forecasted) values are marked by a hat, e.g. \widehat{IV} , while the true (*ex post*) realised volatilities have no hat, i.e. IV . To avoid cluttering the notation, sometimes we omit the hats as well as security- and time-subscripts if there is no risk of misunderstanding.

¹¹⁹ The Spearman's rank-correlation coefficient equals the Pearson's correlation coefficient calculated in the ranks instead of the values. In case of no ties, the co-efficient can also be written as:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)},$$

where $d_i = x_i - y_i$ is the difference between the ranks of the two variables. The coefficient measures the degree of linear dependence between the variables and the coefficient ranged between -1 and $+1$.

¹²⁰ $\rho = \frac{cov(x,y)}{\sigma_x \sigma_y}$

Table 11: Average cross-sectional Spearman rank correlations of idiosyncratic volatilities

The table reports the correlation coefficients between three measures of true volatility and five estimates of expected volatility. $|IRet_{i,t}|$ measures realised volatility as the absolute value of idiosyncratic returns. $IV_{i,t}^{garch}$ infers realised volatility from filtered values from GARCH(1,1) using the full series of monthly returns for each stock. $IV_{i,t}^{egarch}$ infers realised volatility from filtered values from EGARCH(1,1) using the full series of daily returns for each stock.

' IV^{OLS} ' is the volatility of the residual of rolling monthly regressions using 30 to 60 months of data, as available. ' IV_{t-1}^{daily} ' is the idiosyncratic volatility from the daily data in the preceding month calculated using the residuals from Fama-French regressions with the daily data for the preceding month. ' \widehat{IV}_t^{garch} ' is the forecasts from GARCH(1,1) model with SGED innovations estimated using expanding window comprising of at least sixty months of continuous data. ' \widehat{IV}_t^{arma} ' is expected volatility from ARMA(1,1) model fitted on the available series of IV_{t-1}^{daily} .

The table suggests that both *ex ante* (expected) and *ex post* (realised) volatility tend to cluster together based on their calculation frequency. Thus, $IV_{i,t}^{garch}$ calculated from monthly data correlates more closely with \widehat{IV}_t^{ols} and \widehat{IV}_t^{garch} than with the other estimator of realised volatility - $IV_{i,t}^{egarch}$. Likewise, $IV_{i,t}^{egarch}$, calculated from daily returns, correlates more closely with \widehat{IV}_t^{arma} and $\widehat{IV}_{t-1}^{daily}$.

	$IV_{i,t}^{garch}$	$IV_{i,t}^{egarch}$	\widehat{IV}_t^{ols}	\widehat{IV}_t^{garch}	$\widehat{IV}_{t-1}^{daily}$	\widehat{IV}_t^{arma}	m
$ IRet_{i,t} $	0.35	0.37	0.34	0.33	0.27	0.32	0.29
$IV_{i,t}^{garch}$		0.81	0.86	0.90	0.67	0.81	0.76
$IV_{i,t}^{egarch}$			0.78	0.79	0.83	0.91	0.81
\widehat{IV}_t^{ols}				0.92	0.65	0.81	0.81
\widehat{IV}_t^{garch}					0.66	0.82	0.81
$\widehat{IV}_{t-1}^{daily}$						0.85	0.66
\widehat{IV}_t^{arma}							0.85

Source: author's calculations

Spearman's rank correlations for the proxies of the true volatility and the volatility forecasts are reported in Table 11. Consistent with the assertion that squared monthly idiosyncratic returns are a noisy proxy of monthly variance¹²¹, we find that they correlate poorly with the rest of the proxies of true volatility. We find that $|IRet_{i,t}|$ correlation with the other two measures of true (*ex post*) volatility – $IV_{i,t}^{garch}$ and $IV_{i,t}^{egarch}$ – is consequently low and stood at just over 1/3. \widehat{IV}_t^{ols} , \widehat{IV}_t^{garch} and \widehat{IV}_t^{arma} all have fairly close correlation coefficients with monthly returns (correlations in the range 0.32 - 0.34), while the

¹²¹ Since $IRet_{i,t}^2$ is an estimator of variance, $|IRet_{i,t}|$ is an estimator of volatility (standard deviation).

correlation with lagged volatility ($\widehat{IV}_{t-1}^{daily}$, the estimator of Ang et al. (2006)) scores worst (0.27).

The other two estimators of true *ex post* volatility tend to be more tightly correlated (average correlation of 0.81) – certainly a high correlation as such, but an even higher correlation could be expected for two *ex post* estimates of the same unobservable variable – the idiosyncratic volatility. $IV_{i,t}^{garch}$, the *ex post* volatility from GARCH(1,1), in fact correlates more closely with the historical OLS volatility and forecasts from GARCH(1,1) models (mean cross-sectional correlations of 0.86 and 0.90 respectively) than with the other EGARCH(1,1) *ex post* measurement (0.81). One possible explanation could point to the persistence of GARCH volatilities, especially when series lengths are below 10 years – 120 monthly observations, which is reflected in the long half-life¹²² of many of the estimates. The median half-life for the shares in our sample¹²³ is 35.25 months, i.e. almost 3 years. This shows that for many companies GARCH(1,1) with monthly data produces persistent volatility forecasts. This is in contrast with the typical range of persistence of daily volatility series. This latter point is also demonstrated by the wider interquartile distance for EGARCH(1,1) with daily data (7.79 percentage points) compared to the interquartile distance for the true volatilities estimated from GARCH(1,1) with monthly data (6.24 p.p.). Against that backdrop, it should not come as a surprise that the correlation of the EGARCH-measure of true volatilities correlates more closely with the forecasts from ARMA(1,1) and $\widehat{IV}_{t-1}^{daily}$, for the two of which monthly volatilities are inferred from higher-frequency (daily) returns (correlations of 0.91 and 0.83 respectively), than with the other GARCH measure of true volatility that employed only monthly returns (correlation 0.81). Overall, while the two measures of true *ex post* volatilities yield high average cross-sectional correlations, each of these measures tends to favour a forecasting method that was estimated from data with the same frequency as the data underlying the measure of true volatility. A similar pattern is observed for the correlations of volatility forecasts. The forecasts that are based on monthly returns (the OLS historical estimator, and the GARCH(1,1) forecasts) tend to correlate more closely with each other than with the forecasts based on estimates from daily data ($\widehat{IV}_{t-1}^{daily}$

¹²² The half-life (the number of months necessary for volatility to close one half of the spread between current volatility and the unconditional mean of the process) for the GARCH(1,1) $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$, where $\alpha + \beta < 1$ and $\alpha, \beta > 0$ equals $\frac{\ln(1/2)}{\ln(\alpha + \beta)}$.

¹²³ All companies having the same weight irrespective of the number of monthly returns available

and the ARMA(1,1) forecasts). The correlation pattern for the mean-reverting level of volatility is also interesting. Mean-reverting volatility correlates most closely with ARMA(1,1) forecasts. This reflects a material time-variation of the mean-reverting volatilities, which are updated with arrival of new information. Yet mean-reverting volatility correlates much closer with the monthly-frequency volatility forecasts (0.81) than with Ang's historical volatility (0.66), consistent with the hypothesis that the mean-reverting level estimates the equilibrium to which volatility would revert, provided that there was no change in the scale of the distribution of idiosyncratic innovations.

Spearman's rank-correlation coefficient provides an insight into the degree of similarity of the ranking of shares by the different estimators of volatility. However, they provide limited insight into how accurately the *ex ante* forecasts predicted the *ex post* realised returns. In order to avoid the complications of scaling volatilities estimated from higher-frequencies to lower-frequencies, we choose to employ Mincer-Zarnowitz regressions of the form

$$h_{i,t} = a_i + b_i \hat{\sigma}_{i,t}^2 + e_{i,t}, \quad (14)$$

where $h_{i,t}$ was some of the measures of realised variance (squared volatility), and $\hat{\sigma}_{i,t}^2$ is the tested variance forecasting method. A valid test of the model above pre-supposes that $h_{i,t}$ and $\hat{\sigma}_{i,t}^2$ are stationary.

Table 12: Dickey–Fuller tests for *ex ante* and *ex post* volatility estimates

The table reports the results from Dickey-Fuller test statistics for estimators of realised (*ex post*) volatility - $IV_{i,t}^{garch}$ and $IV_{i,t}^{egarch}$ - and expected (*ex ante*) volatilities - \widehat{IV}_t^{ols} , \widehat{IV}_t^{garch} , $\widehat{IV}_{t-1}^{daily}$, \widehat{IV}_t^{arma} , and m . The tests are performed separately for each security, for which there are at least 24 months of data for the respective indicator. Both stationarity of the volatilities (Panel A) and natural logarithm of volatilities (Panel B) are tested. For each series, we have calculated the estimate of the coefficient of the autoregressive term ($\gamma(\cdot)$) and the Dickey-Fuller test statistic ($t(\gamma)$). For each set of tests we report the number of stocks that were tested (N), as well as the mean ('Mean'), median ('Median'), and first and third quartiles ('Q1' and 'Q3') for all calculated values of $\gamma(\cdot)$ and $t(\gamma)$. The share of series for which the unit root null hypothesis is rejected at 1% confidence level is reported in 'UR reject (%)'.

The table demonstrates that the unit root hypothesis is rejected for about half of all realised volatilities $IV_{i,t}^{garch}$ and $IV_{i,t}^{egarch}$. A likely reason is that these estimates are filtered from actual data series and some of the volatility is smoothed out. The most direct measure of volatility that does not involve such smoothing - $\widehat{IV}_{t-1}^{daily}$ - shows that the unit root hypothesis is rejected for 95.37% of all stocks in Panel A and 93.96% of all stocks in Panel B. The forecasts obtained from monthly data - \widehat{IV}_t^{ols} and \widehat{IV}_t^{garch} - demonstrate persistence of volatility forecasts and for these estimators the unit root hypothesis is rejected for a low share of series. Such a mismatch with the persistence properties of the true series also suggests that these forecasts could be inferior predictors of true volatility and may smooth month-on-month changes in volatility.

Variable	N	Mean	Median	Q1	Q3	UR reject (%)
Panel A: $IV_{t+1} - IV_t = \gamma_0 + \gamma IV_t + \varepsilon_t$						
$\gamma(IV_t^{garch})$	4489	-0.06	-0.03	-0.08	-0.00	54.24
$t(\gamma)$		< -10	-3.94	< -10	-2.10	
$\gamma(IV_t^{egarch})$	5507	-0.28	-0.19	-0.40	-0.09	53.50
$t(\gamma)$		-4.12	-3.65	-5.18	-2.55	
$\gamma(\widehat{IV}_t^{OLS})$	5556	-0.03	-0.01	-0.04	-0.00	1.82
$t(\gamma)$		-1.11	-1.13	-1.74	-0.49	
$\gamma(\widehat{IV}_t^{garch})$	4735	-0.14	-0.10	-0.19	-0.05	27.79
$t(\gamma)$		-2.86	-2.55	-3.65	-1.79	
$\gamma(\widehat{IV}_{t-1}^{daily})$	5294	-0.65	-0.64	-0.79	-0.51	95.37
$t(\gamma)$		-7.35	-7.05	-8.97	-5.51	
$\gamma(\widehat{IV}_t^{arma})$	5398	-0.24	-0.15	-0.31	-0.08	43.53
$t(\gamma)$		-3.60	-3.25	-4.42	-2.38	
$\gamma(\widehat{m}_t^{arma})$	5398	-0.09	-0.03	-0.09	-0.01	14.80
$t(\gamma)$		-2.16	-1.80	-2.76	-1.03	

Variable	N	Mean	Median	Q1	Q3	UR reject (%)
Panel B: $\ln(IV_{t+1}) - \ln(IV_t) = \gamma_0 + \gamma \ln(IV_t) + \varepsilon_t$						
$\gamma(IV_t^{garch})$	4489	-0.06	-0.02	-0.07	-0.00	41.68
$t(\gamma)$		< -10	-2.75	< -10	-1.27	
$\gamma(IV_t^{egarch})$	5507	-0.26	-0.18	-0.36	-0.09	48.25
$t(\gamma)$		-3.82	-3.41	-4.90	-2.35	
$\gamma(\widehat{IV}_t^{OLS})$	5556	-0.03	-0.01	-0.04	-0.01	1.48
$t(\gamma)$		-1.07	-1.12	-1.69	-0.48	
$\gamma(\widehat{IV}_t^{garch})$	4735	-0.13	-0.08	-0.16	-0.04	22.94
$t(\gamma)$		-2.61	-2.39	-3.35	-1.67	
$\gamma(\widehat{IV}_{t-1}^{daily})$	5294	-0.61	-0.59	-0.73	-0.46	93.96
$t(\gamma)$		-6.84	-6.59	-8.32	-5.12	
$\gamma(\widehat{IV}_t^{arma})$	5398	-0.22	-0.13	-0.28	-0.06	35.61
$t(\gamma)$		-3.32	-2.94	-4.08	-2.12	
$\gamma(\widehat{m}_t^{arma})$	5398	-0.08	-0.03	-0.09	-0.01	12.88
$t(\gamma)$		-2.01	-1.70	-2.56	-0.93	

Source: author's calculations

In Table 12 we provide results from the Dickey and Fuller (1979) test for unit root.¹²⁴ The null hypothesis is presence of unit root (i.e., $\gamma = 0$). If the process has unit root, then the errors have permanent impact on idiosyncratic volatilities, i.e. they did not die out over time. The table highlights some salient features of the various estimators of *ex ante* and *ex post* volatilities.

We find values of $\gamma(\widehat{IV}_{t-1}^{daily})$ that are quite similar to those reported by Fu (2009). In our sample we reject the null hypothesis in 95% of all shares, compared to approx. 90% in Fu, which may be due to the different composition and time-span of the two samples. On the other hand, the other two measures of *ex post* volatility – the one based on GARCH with monthly data and the other using EGARCH with daily data estimated for the full length of the data series – resulted in volatility measures that reject the unit root hypothesis for only 53-54% of all stocks. At first glance these numbers are surprising as due to the constraints

¹²⁴ The null hypothesis of the test is that $\rho = 0$ in the autoregressive model $y_t = \rho y_{t-1} + \epsilon$.

on the parameters of the GARCH model ($\alpha + \beta < 1$), the GARCH estimates are always with finite variance and mean-reverting. However, if the sum of the two parameters¹²⁵ is sufficiently close to unity, the half-life of the model could increase without bound as the sum approaches one. Indeed, this is consistent with our observations in our sample, where the median half-life is 35.25 months, and the third quartile is 377.8 months. Hence, while in theory the associated model is stationary, in a finite sample of monthly data such mean-reversion could be difficult to estimate and ultimately dependent not on the true properties of the underlying process but rather on the constraints imposed on the parameter estimation. Another possible explanation could be the great parameter uncertainty in the estimation of GARCH parameters due to the scarcity of monthly data. However, such an explanation is contradicted by the similar results obtained by using the EGARCH(1,1) model with aggregated daily data. Or one might hypothesise that there is a volatility trend, e.g. due to the decline of volatility of each security as the company matures and its cash flows become more stable and predictable. However, this does not appear to be the case, as the Dickey-Fuller test with a drift term and trend rejects the unit root hypothesis for an even lower share of companies.

The OLS estimator is at the opposite extreme – for it, the null hypothesis is rejected for less than 2% of all shares. Such behaviour is in line with the interpretation of the OLS estimator as a moving average filter of idiosyncratic shocks. In that case, a unit impulse at the entry of the 60-month moving average filter would be dampened by the filter, but the impulse would have lasting impact on filtered volatility during the 60 months until the shock exits the rolling window. Hence, in short samples we could expect that the unit root null hypothesis would be difficult to reject, which is consistent with what we find in the present sample.

The impact of shocks on the other three volatility forecasts (GARCH forecasts, ARMA forecasts, and ARMA mean-reverting level of volatility) are more difficult to interpret. The GARCH forecasts shared a similar issue as the full-sample *ex post* volatility estimation, namely – the long half-life of the forecasts. However, similarly to the other two measures, significant shocks on volatility could have dual impact: on the one hand, they could result in an increased volatility expectation for the next period (keeping the model parameters fixed). On the other hand, new information also updates the parameter estimates. This latter effect is probably more pronounced for the mean-reverting level of volatility, where the unit

¹²⁵ The bound on the sum of the two parameter, $\alpha + \beta < 1$, refers only to GARCH(1,1) model.

root hypothesis is rejected for less than 15% of all securities.

Overall, we note that the various volatility estimates have markedly different time-series properties, even though they all tended to produce not too dissimilar rankings of securities in the cross-section. At the one end is the quite static OLS estimator, where large shocks produced long-lasting impact on volatility estimates. On the other extreme is the measure used by Ang et al. (2006) ($\widehat{IV}_{t-1}^{daily}$), where the volatility in each month is determined solely by the returns in that period. Somewhere in the middle is the (E)GARCH estimator where month- t volatilities are also affected by prior-months' returns, so that the variability of filtered or predicted volatilities is somewhat smoothed at the cost of more durable impact of volatility shocks on estimated volatilities.

We now move to the question of which idiosyncratic risk estimators produce more accurate estimates of future (one-period ahead) volatilities. As discussed in the methodology chapter, that approach could not be properly addressed using the loss function approach because some of the *ex ante* and *ex post* measures are based on daily data that are subsequently scaled to monthly frequency, and the scaling method could obscure comparisons with those measures that are estimated directly from monthly data. Thus, the loss function approach that is used, for example, by Spiegel and Wang (2005) cannot be applied in our setting. Instead, as discussed in the chapter on methodology, we follow Mincer and Zarnowitz (1969) and Pagan and Schwert (1990) and we compare volatility forecasts using the standard predictive regressions.

Table 13: Predictive accuracy from Mincer-Zarnowitz regressions

The table reports aggregated results from comparison of the accuracy of volatility forecasts using Mincer-Zarnowitz regressions, where $h_{i,t}$ is some measure of realised variance (either monthly volatility filtered from EGARCH(1,1) with daily returns, or squared monthly idiosyncratic return), and $\hat{\sigma}_{i,t}^2$ is the tested variance forecasting method - \widehat{IV}_t^{ols} , \widehat{IV}_t^{garch} , $\widehat{IV}_{t-1}^{daily}$, \widehat{IV}_t^{arma} . Panels A, B and C test three alternative specifications that are run for each available security with more than 24 months of data. Parameter estimates for each specification are obtained using two methods – ordinary least squares (OLS), or the more robust median regression. For each regression we calculate a measure of goodness of fit - either R^2 (for OLS regressions), or its analogue for quantile regression - $R^1(0.5)$, proposed by Koenker and Machado (1999). For each specification we report the mean ('mean') and median ('median') values of the set of goodness-of-fit measures (R^2 or $R^1(0.5)$). '# of firms' reports the number of firms, for which the respective Mincer-Zarnowitz regression was estimated. '% of pos.' reports the share of cases where the goodness-of-fit measure for the respective case was superior compared to the corresponding goodness-of-fit measure with \widehat{IV}_t^{arma} as predictor variable. The results show that \widehat{IV}^{arma} is vastly superior to the other volatility forecasts, especially when true volatility is filtered from EGARCH, followed by IV_{t-1}^{daily} . \widehat{IV}_t^{ols} and \widehat{IV}_t^{garch} on the other hand outperformed \widehat{IV}_t^{arma} is less than 13% of all cases, when realised volatility was based on EGARCH. Thus, a simple binomial test implies that \widehat{IV}_t^{arma} is a superior predictor of volatility against all other tested estimators and in all specifications.

Variance	h_t from EGARCH				$h_t = IRet_t^2$			
forecast	mean	median	# of firms	% of pos.	mean	median	# of firms	% of pos.
	Panel A: $h_{i,t} = a_i + b_i \hat{\sigma}_{i,t}^2$							
	<i>OLS regression – R^2</i>							
\widehat{IV}_t^{arma}	41.07	42.96	5562	–	4.70	2.34	5591	–
\widehat{IV}_t^{ols}	12.38	6.14	5562	13.90	2.60	1.00	5591	36.02
$\widehat{IV}_{t-1}^{daily}$	37.87	39.71	5374	32.71	4.17	1.38	5394	33.30
\widehat{IV}_t^{garch}	16.49	10.50	4833	11.75	2.50	1.12	4857	34.90

Variance	h_t from EGARCH				$h_t = IRet_t^2$			
forecast	mean	median	# of firms	% of pos.	mean	median	# of firms	% of pos.
	<i>Median regression – $R^1(0.5)$</i>							
\widehat{IV}_t^{arma}	29.09	29.06	5564	–	1.67	0.88	5591	–
\widehat{IV}_t^{ols}	8.09	4.23	5564	10.73	1.09	0.48	5591	36.56
$\widehat{IV}_{t-1}^{daily}$	25.15	25.68	5376	29.63	1.64	0.77	5394	40.01
\widehat{IV}_t^{garch}	10.68	7.20	4835	9.06	1.12	0.56	4857	38.07
	Panel B: $\sqrt{h_{i,t}} = a_i + b_i \hat{\sigma}_{i,t}$							
	<i>OLS regression – R^2</i>							
\widehat{IV}_t^{arma}	45.06	48.40	5564	–	5.43	3.24	5591	–
\widehat{IV}_t^{ols}	13.77	7.17	5564	12.29	3.13	1.41	5591	33.11
$\widehat{IV}_{t-1}^{daily}$	44.82	48.90	5376	39.55	4.73	2.66	5394	34.80
\widehat{IV}_t^{garch}	18.24	12.71	4835	9.82	3.25	1.63	4857	32.55
	<i>Median regression – $R^1(0.5)$</i>							
\widehat{IV}_t^{arma}	31.06	31.41	5564	–	2.55	1.53	5591	–
\widehat{IV}_t^{ols}	8.84	4.79	5564	10.32	1.82	0.85	5591	36.38
$\widehat{IV}_{t-1}^{daily}$	29.48	30.97	5376	37.07	2.30	1.38	5394	37.91
\widehat{IV}_t^{garch}	11.47	7.83	4835	8.17	1.82	0.95	4857	36.52

Variance	h_t from EGARCH				$h_t = IRet_t^2$			
forecast	mean	median	# of firms	% of pos.	mean	median	# of firms	% of pos.
	Panel C: $\ln(h_{i,t}^2) = a_i + b_i \ln(\hat{\sigma}_{i,t}^2)$							
	<i>OLS regression – R²</i>							
\widehat{IV}_t^{arma}	47.02	50.85	5539	–	3.63	2.13	5564	–
\widehat{IV}_t^{ols}	14.63	7.96	5564	11.36	2.57	1.13	5591	35.83
$\widehat{IV}_{t-1}^{daily}$	47.72	52.20	5376	42.06	3.28	1.93	5394	37.60
\widehat{IV}_t^{garch}	19.00	13.46	4835	9.02	2.58	1.26	4857	36.11
	<i>Median regression – R¹(0.5)</i>							
\widehat{IV}_t^{arma}	31.89	32.23	5539	–	2.42	1.47	5564	–
\widehat{IV}_t^{ols}	9.24	5.06	5564	10.26	1.79	0.83	5591	36.04
$\widehat{IV}_{t-1}^{daily}$	30.75	32.25	5376	39.58	2.18	1.32	5394	37.30
\widehat{IV}_t^{garch}	11.78	7.97	4835	8.31	1.77	0.91	4857	36.09

Source: author's calculations

The Mincer-Zarnowitz regressions are estimated individually for each security. We require at least 24 months of pairwise-complete data. The mean and median R^2 's are reported in Table 13. The first important result from that exercise is the poor performance of the forecasts from GARCH(1,1) in explaining either of the two proxies of the true volatility. With either of the two measures of true volatility employed in this study, the results for \widehat{IV}_t^{garch} are closer to that of the variance of the residual from the OLS regression (IV^{OLS}) than to the forecasts from ARMA(1,1) (\widehat{IV}^{arma}) or the estimator used by Ang et al. (2006) and Ang et al. (2009) (IV_{t-1}^{daily}). Under either of the two proxies of true volatility, the estimator used by Ang et al for period- t volatility in fact outperforms the forecast from GARCH(1,1).

The magnitude of the differences found in the comparison is indicative of the significance of the difference. At each row we provide the percentage share of securities for each the R^2 (for OLS regression) or R^1 (for median regression) that have outperformed the respective measure (R^2 or R^1) for ARMA(1,1); that percentage is provided in column “% of pos.”. In case the respective measure outperforms ARMA(1,1) we would see that share exceeding 50%. The table does not include results from statistical tests of the observed differences, however the significance of the results can be gauged from the small number of positive differences. Indeed, we could consider the numbers of positive and negative differences as the outcome of Bernoulli trials and calculate the significance using binomial test.¹²⁶ For an average of about 5,000 securities, the two-sided 99% confidence interval would be between 0.4817 and 0.5183. In all cases in Table 13 the percentages are materially outside that interval, hence we should conclude that with high confidence the observed under-performance of all models relative to ARMA(1,1) is not mere happenstance.

The second important feature is the low values of R^2 's when squared idiosyncratic returns are used as dependent variable. The low value of R^2 in Mincer-Zarnowitz regressions is well documented. The previously referred study of Andersen and Bollerslev (1998) proves that the reason for the low R^2 s is not a failure of the GARCH models but rather the fact that squared returns are a noisy estimator of volatility. In particular they show that the R^2 s for GARCH(1,1) models are bounded above by $1/\kappa$, where κ is the kurtosis of the underlying

¹²⁶ If two competing models have the same predictive performance as measured by R^2 , then we would expect that the difference of the R^2 s from the competing Mincer-Zarnowitz regressions would be positive in half of the cases, with a number of positive or negative differences following binomial distribution. The benefit of that test is that it is exact and does not require knowledge of the distribution of the differences of R^2 , including no assumption of symmetry of the distribution of the differences.

noise distribution. In this case, however, the R^2 s are particularly low. A plausible reason is that the four factors in the mean equation account for much of the predictable variability, and the remaining idiosyncratic component is unpredictable and not persistent (so that past values are not useful for forecasting). Alternatively, it is possible that EGARCH significantly improves forecasts compared to the simpler GARCH(1,1) model. However, the sharp contrast with the results where true volatility is proxied by the average volatility from the EGARCH(1,1) model suggested that GARCH forecasts in the present context are under-performing materially the estimates from the ARMA(1,1) and the lagged volatility IV_{t-1}^{daily} .

Considering the previous observation that the parameters of a typical GARCH model in our sample implies a half-life of about 3 years, and observing the very similar predictive performance of GARCH(1,1) with the 24-60 months rolling average (IV^{OLS}), a likely explanation of the results emerges: the low number of months available for GARCH estimation and the inherent noise contained in monthly idiosyncratic returns results in GARCH models with monthly data forecasting some mean, longer-term component of volatility and failing to track with satisfactory accuracy the month-on-month variation of volatilities.

Overall, the results in Table 13 suggested that forecasts of monthly volatilities from ARMA(1,1) outperforms the remaining estimators studied here. Moreover, the historical measure of Ang et al (IV_{t-1}^{daily}) outperforms GARCH estimates obtained from monthly returns; in fact, for many series the latter are closely correlated to the simple OLS volatilities. This suggests that the significance of GARCH forecasts in explaining the cross-section of returns that is observed by Fu might be misleading and in fact the direction of correlation could be negative, as argued by Ang et al. In the following section we shall examine this issue and compare and contrast the power of the alternative volatility forecasts to explain the cross-section of returns.

4.3. Idiosyncratic volatility and the cross-section of stock returns: results from Fama–Macbeth cross-sectional regressions

We evaluate the significance of idiosyncratic volatility in explaining the cross-section of market returns using the Fama and MacBeth (1973) cross-sectional regressions. At each

calendar date in our sample for which there is available information on capitalisation (*Cap*), CAPM beta (*Beta*), and positive book-to-market ratio (*B/M*), we estimate the following cross-sectional regressions

$$R_{it} = \gamma_{0t} + \gamma_{1t}\beta_{1it} + \dots + \gamma_{mt}\beta_{mit} + \varepsilon_{it},$$

where β_{kit} is the loading of stock $i = 1 \dots N$ on factor $k = 1 \dots K$ at time $t = 1 \dots T$, and γ_{kt} are the parameters to be estimated. The estimates $\hat{\gamma}_{kt}$ are treated as random variables and the mean of the estimates from the individual cross-sectional regressions is used as an estimator of the true value ($\hat{\gamma}_k$), i.e. $\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^T \hat{\gamma}_{kt}$. The standard error of the estimates is calculated using the Newey and West (1987) autocorrelation-corrected coefficients with a lag of 4.

The method of Fama–Macbeth could be employed using portfolios or individual securities as base assets in the test. Apart from market betas, the explanatory factors in our sample, including capitalisation, B/M, idiosyncratic volatility (a second moment), and to a somewhat lesser extent, the liquidity spread based on auto-covariances, could be estimated with reasonable accuracy. Nevertheless, in order to mitigate the impact of extreme values of the explanatory factors on the cross-sectional regressions, each month the explanatory variables are censored at the 0.005 and 0.995 quantiles.

Securities with low prices often have greater noise in their returns, which is related to the minimum step by which prices move, causing abrupt discrete changes in share prices. To mitigate this concern, studies usually remove from the sample securities with an unadjusted price below some selected threshold value. We opt for a low threshold of \$1 for most regressions. Some studies¹²⁷ opt for a higher threshold of \$10. The choice of threshold could affect conclusions. For example, Brandt et al. (2010: 881) document that the level of institutional ownership and market capitalisation are significantly lower for low-priced stocks (defined as those in price deciles 1 through 3). The low-priced stocks typically have a price below \$10, market capitalisation below \$100 million, and institutional ownership below 10 per cent. Therefore, setting a threshold of \$10 would discard proportionately more securities with high retail ownership, where investors are typically less diversified and thus more sensitive to idiosyncratic risk. Hence, we employ a baseline threshold of \$1, but key relationships were also tested for robustness against higher thresholds. The threshold (filter) applied to each regression model is reported in the tables.

There are a number of control variables in the Fama-Macbeth cross-sectional

¹²⁷ e.g. Bali and Cakici (2008)

regressions. In order to facilitate presentation, we split the results into two tables: one containing the results for OLS and GARCH volatilities (Table 14), and another one – for Ang et al and ARMA(1,1) (Table 15).

The results from the Fama–MacBeth regressions with OLS and GARCH volatilities are presented in Table 14 on p. 148. Models 1 through 5 list the cross-sectional estimates for the CAPM specification as well as the CAPM augmented with the three characteristics used as a baseline explanation of the cross-section of returns. In particular, model 1 corresponds to the standard CAPM model. The significance of the CAPM beta varies in the different specifications, but in view of its strong theoretical justification, we keep it in all specifications even when insignificant, in order to avoid omitted variable bias in our results. Model 2 corresponds to the CAPM with the two Fama and French factors. Model 3 adds to the three-factor Fama–French specification the return momentum, calculated as the return between $(t - 7)$ and $(t - 2)$. Model 4 further adds the liquidity spread, and model 5 adds previous month return in order to account for return reversal. The results for models 1 through 5 are consistent with the existing body of literature and confirmed that the two Fama–French factors, liquidity, momentum, and lagged return are all significant predictors of the cross-section, while beta is an unreliable predictor of the cross-section of returns.

Specifications 6 through 8 introduce the monthly idiosyncratic volatility from the OLS regressions. In all three specifications we find that IV^{ols} is a significant predictor of the cross-section. Moreover, its magnitude and significance do not change substantially between the three compared models, suggesting that the significance of IV^{ols} is not a result of its correlation with another predictor.

Idiosyncratic volatilities usually exhibit significant skewness (cf Table 6 on p. 117). In order to examine whether the significance of idiosyncratic volatilities is a result of their skewness, in models 9 through 11 we also add the natural logarithm of the idiosyncratic volatility, $\ln(IV^{ols})$, to the model specification. As seen in Table 6, the log-transform of volatilities substantially reduces the skewness of the variable and thus mitigates the concerns that the significance of the regression coefficient is an artefact purely of few high-volatility securities. The results support the significance of IV^{ols} as a predictor of the cross-section, as the magnitude of the log of volatilities remains numerically stable and significant in the tested specifications.

In specifications 12 through 17 we repeat the same exercise, but instead of the OLS volatilities, we employ the one-month-ahead forecasts from the GARCH(1,1) model. We

previously observed that these forecasts have long half-lives and correlated with the OLS volatilities. The long half-lives suggest that shocks on volatilities have long-lasting, nearly-permanent effect, i.e. in finite samples volatilities could behave similarly to random walk. On the other hand, the high correlation with the OLS volatilities suggests that such shocks that result in material quasi-permanent change of volatility are infrequent and volatility of each share tends to remain in a narrow band for relatively longer periods. This is consistent with the experience of the present author from the inspection of individual series of some high-profile stocks when volatility forecasts are obtained from GARCH(1,1) models estimated with monthly data. The latter qualification ('with monthly data') suggests that for short periods (less than a month up to a couple of months) the volatility of individual shares may vary materially but quickly reverts to some average level, and hence there are not enough months of exceptionally high- or low-magnitude shocks that would result in the upward or downward revision of the volatility forecasts from GARCH(1,1).

Having regard for the foregoing discussion, we should expect that \widehat{IV}_t^{garch} would be a significant predictor of the cross-section, and that its sign and magnitude would be similar to that of IV^{ols} . Indeed, in general the results in Table 14 are consistent with that expectation. Both volatilities and their log-transform remain significant predictors of the cross-section, although the t statistics are palpably lower compared with the significance of the corresponding specifications involving IV^{ols} . Increasing the threshold of small-value stocks from US\$ 1 to US\$ 10 decreases the significance of the idiosyncratic volatilities; in some specifications, esp. 13a–15a, idiosyncratic volatilities are no longer significant predictors of the cross-section. The change of significance depending on the threshold is consistent with the explanation of different investor profiles for securities with smaller prices, which are preferred by individual investors, who are also known for being less diversified, and hence those securities earn a higher risk-premium for idiosyncratic volatility. On the other hand, our results are consistent with those obtained by Bali and Cakici (2008) who find that the correlation between idiosyncratic volatility and returns is not robust and changes across specifications.

Finally, we note that even when idiosyncratic volatility is an insignificant predictor, the point estimate does not change sign, suggesting that longer series or higher frequencies through the associated larger sample sizes might offer more conclusive evidence on idiosyncratic volatility significance.

Table 14: Fama–Macbeth cross-sectional regressions with OLS and GARCH forecasts

The table reports results from Fama-Macbeth cross-sectional regressions. Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll’s estimator (*Roll*). Tested variables are the OLS measure of idiosyncratic volatility (IV^{ols}) and from GARCH(1,1) (\widehat{IV}_t^{garch}), and the natural logarithms of these measures. ‘ R^2 ’ reports the averaged R-squared statistics from the cross-sectional regressions. Column ‘Filter’ indicates the subsample used for estimating the cross-sectional regressions: ‘all’ in case the entire sample is used; ‘UP>10’ indicates specifications estimated using only securities with unadjusted price of over USD 10. Newey-West t-statistics are reported in parenthesis. The results suggest that IV^{ols} and \widehat{IV}_t^{garch} are robust statistically significant predictors of the cross-section of returns, with sign and magnitude stable across specifications. The regression coefficient tends to be somewhat lower in the specifications that involve only stocks with unadjusted price above USD 10, suggesting that idiosyncratic risk premium is likely higher among stocks with lower unadjusted price – a segment preferred by individuals due to lower transaction costs.

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	IV^{ols}	$\ln(IV^{ols})$	\widehat{IV}_t^{garch}	$\ln(\widehat{IV}_t^{garch})$	R^2
1	all	0.84 (-2.37)										1.81
2	all	0.69 (2.21)	-0.10 (-2.50)	0.79 (7.16)								3.45
3	all	0.63 (2.08)	-0.09 (-2.48)	0.90 (9.51)	0.01 (4.83)							4.16
4	all	0.48 (1.83)	-0.01 (-0.45)	0.93 (10.83)	0.01 (4.68)	0.05 (2.42)						4.66
5	all	0.42 (1.57)	0.00 (-0.13)	0.85 (10.23)	0.01 (4.05)	0.06 (2.57)	-0.04 (-8.94)					5.26
6	all	0.41 (1.62)						0.05 (2.90)				2.65

#	Filter	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7,-2)$	$Roll$	Ret_{t-1}	IV^{ols}	$\ln(IV^{ols})$	\widehat{IV}_t^{garch}	$\ln(\widehat{IV}_t^{garch})$	R^2
6a	UP>10	0.31 (1.41)						0.02 (1.12)				3.48
7	all	0.30 (1.24)	-0.02 (-0.51)	0.90 (9.28)				0.06 (3.98)				4.02
7a	UP>10	0.27 (1.24)	-0.01 (-0.24)	0.58 (6.59)				0.04 (2.19)				4.82
8	all	0.24 (1.05)	0.02 (0.53)	0.9 (11.27)	0.01 (4.01)	0.04 (1.87)	-0.04 (-9.15)	0.04 (3.45)				5.52
8a	UP>10	0.24 (1.13)	-0.02 (-0.49)	0.62 (8.18)	0.01 (3.96)	-0.03 (-1.17)	-0.03 (-6.36)	0.03 (2.66)				6.62
9	all	0.24 (1.06)							0.67 (3.05)			2.71
9a	UP>10	0.19 (0.95)							0.31 (1.45)			3.49
10	all	0.16 (0.75)	0.00 (-0.06)	0.90 (9.19)					0.81 (3.82)			4.12
10a	UP>10	0.15 (0.74)	0.01 (0.16)	0.59 (6.46)					0.51 (2.51)			4.88
11	all	0.14 (0.66)	0.03 (0.80)	0.9 (11.21)	0.01 (4.14)	0.03 (1.75)	-0.04 (-9.18)		0.56 (3.44)			5.62

#	Filter	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	IV^{ols}	$\ln(IV^{ols})$	\widehat{IV}_t^{garch}	$\ln(\widehat{IV}_t^{garch})$	R^2
11a	UP>10	0.14 (0.68)	-0.01 (-0.23)	0.63 (8.14)	0.01 (3.98)	-0.04 (-1.66)	-0.03 (-6.33)		0.5 (3.47)			6.68
11b	UP>10	0.18 (0.93)	-0.01 (-0.17)	0.68 (8.61)	0.01 (4.78)	-0.06 (-2.38)			0.50 (3.42)			6.08
12	all	0.46 (1.83)								0.03 (2.01)		2.79
12a	UP>10	0.40 (1.79)								0.00 (0.22)		3.51
13	all	0.39 (1.60)	-0.02 (-0.74)	0.84 (8.73)						0.04 (2.34)		4.21
13a	UP>10	0.38 (1.66)	-0.02 (-0.56)	0.53 (5.96)						0.01 (0.81)		4.92
14	all	0.30 (1.27)	0.01 (0.43)	0.85 (10.66)	0.01 (3.61)	0.04 (2.10)	-0.04 (-9.82)			0.02 (1.47)		5.74
14a	UP>10	0.33 (1.55)	-0.02 (-0.63)	0.58 (7.79)	0.01 (3.38)	-0.03 (-0.94)	-0.03 (-6.70)			0.01 (1.01)		6.76
15	all	0.32 (1.37)									0.56 (2.57)	2.83
15a	UP>10	0.30 (1.42)									0.17 (0.83)	3.5

#	Filter	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	IV^{ols}	$\ln(IV^{ols})$	\widehat{IV}_t^{garch}	$\ln(\widehat{IV}_t^{garch})$	R^2
16	all	0.29 (1.28)	-0.01 (-0.31)	0.84 (8.71)							0.59 (2.80)	4.28
16a	UP>10	0.28 (1.30)	-0.01 (-0.20)	0.55 (6.02)							0.30 (1.52)	4.93
17	all	0.24 (1.09)	0.02 (0.57)	0.85 (10.62)	0.01 (3.70)	0.04 (1.83)	-0.04 (-9.84)				0.36 (2.17)	5.81
17a	UP>10	0.26 (1.30)	-0.01 (-0.46)	0.59 (7.84)	0.01 (3.42)	-0.04 (-1.55)	-0.03 (-6.74)				0.32 (2.20)	6.79
17b	UP>10	0.31 (1.51)	-0.01 (-0.46)	0.63 (8.26)	0.01 (4.14)	-0.05 (-2.00)					0.28 (1.94)	6.18

Source: author's calculations

The results in Table 14 provide some support for the hypothesis that idiosyncratic volatility is a significant predictor of the cross-section. Nevertheless, we should also recall that in a preceding section we argued that IV^{ols} and \widehat{IV}_t^{garch} are inferior predictors of expected volatility compared to IV_{t-1} and \widehat{IV}_t^{arma} . Against that backdrop it is important to examine how those arguably more accurate predictors of next-period volatility predict the cross-section of returns. If next-period idiosyncratic volatility is a true explanatory characteristic, then we should expect superior significance of those superior estimators. The results from those tests are presented in Table 15.

The results in Table 15 come as a surprise in view of the encouraging results obtained with IV^{ols} and \widehat{IV}_t^{garch} . Specifications 1 through 3 explore the correlation between the last-period idiosyncratic volatility and returns. Consistent with the findings of Ang et al. (2009, 2006), in specifications 2 and 3 we find a statistically-significant negative correlation between returns and past-month idiosyncratic volatility, IV_{t-1} . The result remains true when we control for the unadjusted price of the securities in our portfolio, i.e. when we allow only securities with unadjusted price above US\$ 10 in the test sample. In specifications 4, 5 and 6 we employ log-transform of idiosyncratic volatility in order to reduce the skewness of the explanatory variable. This also allows us to reduce the impact on the regression coefficient of the high-volatility securities in each month; Fu (2009) argues that the reversals of returns of the highest-quintile stocks are a probable reason for the significant negative correlation observed in the data. We find that the negative correlation documented by Ang et al. is not robust in our sample. Only specification 6 offers a strong negative correlation, yet that evidence is substantially weakened by the results reported in 6a, which documents no significant correlation. Thus, the evidence appears to be consistent with the results of Fu and suggests that much of the explanatory power of IV_{t-1} is placed in the right tail of cross-sectional volatility distribution and that the negative correlation is detected primarily among high-volatility, low-price stocks.

Fu points out that last-period volatility is not a forward-looking estimator of next-period volatility, unless volatilities follow random walk, a hypothesis that he rejects for the overwhelming majority of his sample, and which we confirm in our sample as well. We address this deficiency of IV_{t-1} by using \widehat{IV}_t^{arma} to estimate next-period expected volatility. Furthermore, in the preceding section we reasoned that that measure (\widehat{IV}_t^{arma}) offered the highest predictive accuracy among the four estimators studied in this chapter.

Therefore, we should expect that \widehat{IV}_t^{arma} would offer the most relevant evidence on the correlation between next-month volatility and returns. In specifications 7 through 12 we explore whether idiosyncratic volatilities from our best predictor of next-period monthly volatility explain the cross-section of next-month realised returns. In those specifications we find that idiosyncratic volatilities are not significantly correlated with returns. This is evidenced both by the small values of the t -statistic in most regressions, as well as the changes of sign of the point estimates across specifications.

Table 15: Fama–Macbeth cross-sectional regressions with historical and ARMA forecasts

The table reports results from Fama-Macbeth cross-sectional regressions. Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll's estimator (*Roll*). Tested variables are Ang's previous-month idiosyncratic volatility (\widehat{IV}_{t-1}) and ARMA(1,1) (\widehat{IV}_t^{arma}), and the natural logarithms of these measures. R^2 reports the averaged R-squared statistics from the cross-sectional regressions. Column 'Filter' indicates the subsample used for estimating the cross-sectional regressions: 'all' in case the entire sample is used; 'UP>10' indicates specifications estimated using only securities with unadjusted price of over USD 10. Newey-West t-statistics are reported in parenthesis.

The results suggest that \widehat{IV}_{t-1} and \widehat{IV}_t^{arma} are not robust predictors of the cross-section of returns. Cross-sectional regressions involving \widehat{IV}_{t-1} confirm the findings of previous studies of negation correlation with returns. However, using log-transformed volatility shows that the negative coefficients are result principally to the skewness of the explanatory variable and the negative correlations does not generalise to the whole sample. \widehat{IV}_t^{arma} is not significant predictor either and its coefficient changes sign and significance between specifications even though previous analyses suggest that it is the best predictor of volatility among the examined ones.

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	\widehat{IV}_{t-1}	$\ln(\widehat{IV}_{t-1})$	\widehat{IV}_t^{arma}	$\ln(\widehat{IV}_t^{arma})$	R^2
1	all	0.64 (2.12)						0.00 (0.12)				2.75
1a	UP>10	0.46 (1.68)						-0.01 (-1.27)				3.17
2	all	0.58 (2.03)	-0.13 (-3.45)	0.73 (6.49)				-0.02 (-1.98)				4.20
2a	UP>10	0.46 (1.71)	-0.07 (-2.17)	0.5 (4.66)				-0.02 (-1.82)				4.64
3	all	0.33 (1.35)	-0.05 (-1.37)	0.81 (9.10)	0.01 (3.91)	0.09 (3.18)	-0.03 (-7.23)	-0.04 (-5.35)				5.93
3a	UP>10	0.33 (1.50)	-0.05 (-1.56)	0.58 (7.09)	0.01 (3.85)	0.01 (0.25)	-0.02 (-5.18)	-0.02 (-2.32)				6.77

#	Filter	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	\bar{IV}_{t-1}	$\ln(\bar{IV}_{t-1})$	\bar{IV}_t^{arma}	$\ln(\bar{IV}_t^{arma})$	R^2
4	all	0.53 (1.88)							0.22 (1.49)			2.76
4a	UP>10	0.41 (1.56)							-0.01 (-0.12)			3.18
5	all	0.53 (1.95)	-0.11 (-2.93)	0.74 (6.52)					0.00 (0.00)			4.22
5a	UP>10	0.42 (1.64)	-0.07 (-1.91)	0.51 (4.73)					-0.05 (-0.43)			4.66
6	all	0.35 (1.46)	-0.05 (-1.34)	0.80 (9.03)	0.01 (4.19)	0.06 (2.32)	-0.03 (7.51)		-0.2 (-2.58)			5.93
6a	UP>10	0.33 (1.51)	-0.05 (-1.53)	0.58 (7.07)	0.01 (3.98)	-0.01 (-0.25)	-0.02 (-5.34)		-0.04 (-0.51)			6.75
7	all	0.40 (1.46)								0.03 (2.21)		2.91
7a	UP>10	0.39 (1.64)								0.00 (-0.23)		3.55
8	all	0.51 (1.94)	-0.01 (-0.20)	0.83 (8.17)						0.02 (1.22)		4.26
8a	UP>10	0.44 (1.88)	-0.03 (-0.85)	0.54 (5.81)						-0.01 (-0.34)		4.99

#	Filter	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	\bar{IV}_{t-1}	$\ln(\bar{IV}_{t-1})$	\bar{IV}_t^{arma}	$\ln(\bar{IV}_t^{arma})$	R^2
9	all	0.40 (1.60)	0.01 (0.22)	0.85 (10.10)	0.01 (3.69)	0.05 (2.19)	-0.04 (-8.83)			-0.01 (-0.63)		5.75
9a	UP>10	0.37 (1.69)	-0.03 (-0.83)	0.59 (7.59)	0.01 (3.68)	-0.02 (-0.68)	-0.03 (-6.03)			0.00 (-0.08)		6.77
10	all	0.30 (1.21)									0.54 (2.53)	2.90
10a	UP>10	0.33 (1.44)									0.09 (0.41)	3.51
11	all	0.44 (1.85)	0.00 (0.00)	0.83 (8.15)							0.39 (1.69)	4.32
11a	UP>10	0.39 (1.74)	-0.02 (-0.54)	0.55 (5.73)							0.09 (0.43)	4.99
12	all	0.39 (1.65)	0.01 (0.29)	0.85 (10.09)	0.01 (3.89)	0.05 (2.00)	-0.04 (-8.87)				0.07 (0.45)	5.83
12a	UP>10	0.33 (1.55)	-0.02 (-0.47)	0.60 (7.62)	0.01 (3.75)	-0.05 (-1.69)	-0.03 (-6.07)				0.28 (1.83)	6.80

Source: author's calculations

In this section we confirmed the conflicting evidence concerning the sign of the correlation between idiosyncratic volatilities and returns. We found that the correlation was fairly robust when we forecast idiosyncratic volatility through the residual of monthly OLS regressions or through GARCH(1,1) forecasts, yet the sign turned to negative or became insignificant when we employed last-month volatility or volatility forecasts from ARMA model. This result was puzzling given the results from the previous section, where we found that the latter two estimators of volatility were better forecasts for the unobservable true volatility compared to the former two. If the more accurate forecasts of volatilities were not significantly linked to returns, then there was some other information in the former estimators (\widehat{IV}_t^{garch} and \widehat{IV}_t^{ols}) that was helpful in explaining the cross-section and improving the accuracy of predicting volatility would offer limited contribution to explaining the cross-section of returns. We dedicate the next section to exploring this problem.

4.4. Mean-reverting level of volatility

The last section posed an interesting problem: we found that the more accurate prediction of next-period volatilities did not result in better forecasts of cross-sectional returns as ARMA forecasts generally were not significantly correlated with returns. The standard CAPM model is of limited help in this case, as it assumes that volatilities are fixed. Fixed volatilities are a reasonable assumption that allows examination of equilibrium prices and returns in the CAPM model, which emphasises the more elusive beta, whereas variance could be estimated with higher precisions; in reality, however, volatilities do vary from period to period.¹²⁸ Furthermore, our ability to predict changes in future volatilities is inherently limited: the GARCH models essentially assume that next-month volatility can be predicted from its values in the preceding months, but rarely and with limited success are able to incorporate any fundamental explanatory variables, be that macroeconomic indicators or security characteristics. The simple GARCH(1,1) model offered the most transparent

¹²⁸ A particular difficulty in incorporating stochastic volatility in the CAPM model is that the model assumes complete markets. If volatility of stocks changes randomly, then there may not exist an unique risk-neutral measure, and the same contingent claim would have different values under different risk-neutral measures; for example of conditions for market completeness under stochastic volatility see Davis (2004).

demonstration of this fact as next-period volatility is just the weighted sum of the previous-period forecast and the previous-period squared return (a measure of the realised volatility). Thus volatility forecasts are formed from past forecasts that are continuously updated with a proxy of the last-period realised volatility. The update rule could be symmetric as in the simple GARCH model, but could also be asymmetric, as is the case with the EGARCH model. Nonetheless, fundamental variables (e.g. interest rates, economic growth, credit risk spreads, etc) are rarely present in such specifications; essentially, volatility forecasts are function solely of the past history of volatility forecasts and the distribution of realised returns, especially its variance and skewness. Therefore, if the GARCH forecasts and the OLS forecasts contain information about expected returns that is not present in the ARMA forecasts or lagged (IV_{t-1}) volatilities, it is reasonable to conjecture that such useful information was contained in some other characteristic of the volatility process. Such a characteristic could be the mean-reverting level of volatility from the ARMA model. Indeed, if volatility was random walk, then such mean reversion would not exist. However, as we confirmed in the previous section, the unit root hypothesis for IV_{t-1} was rejected for the overwhelming majority of securities. The ARMA model produced best forecasts among the examined models, and these forecasts were based on the history of IV_{t-1} , which was found to be stationary. We argued that the superior predictive power of ARMA was due to its reliance on a less noisy proxy of realised volatility (average daily volatility scaled up to monthly frequency) to update the next period forecast compared to the GARCH model with monthly data that used squared monthly return as a proxy of realised volatility in the respective month. Therefore, we also examined the predictive performance of the mean-reverting level of volatility estimated from the ARMA model for explaining the cross-section of returns.¹²⁹ In the case of ARMA(1,1) model, $X_t = \mu + \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$, and the mean-reverting level was the natural explanatory variable. Indeed, we estimated \widehat{IV}_t^{arma} using the history of IV_{t-1} and neither \widehat{IV}_t^{arma} (the conditional expectation), nor IV_{t-1} offered a significant prediction of the cross-section, which leaves the unconditional mean of the process as the natural alternative explanatory variable ($E(X_t) = \frac{\mu}{1-\phi} = m$).

¹²⁹ Let X_t be a stationary stochastic process that has ARMA(1,1) representation, $X_t = \mu + \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$. Then:

$$E(X_t) = \frac{\mu}{1-\phi} = m,$$

$$var(X_t) = \frac{1+2\phi\theta+\theta^2}{1-\phi^2} \sigma_\varepsilon^2,$$

$$cov(X_t, X_{t-1}) = \frac{(\phi+\theta)(1+\phi\theta)}{1-\phi^2} \sigma_\varepsilon^2.$$

The results from those tests are summarised in Table 16. The table shows that the mean-reverting level of volatility is indeed a robust predictor of the cross-section. The point estimates remained stable and significant when both the mean-reverting level and its log are used as explanatory variables. Moreover, the estimated slopes are very similar to those found in OLS and GARCH(1,1) tests reported previously, which supports the explanation that those measures had significant predictive performance due to their correlation with the mean-reverting level, rather than next-period volatility. Filtering only securities with unadjusted price over 10 US dollars did not materially affect either the slopes, or their level of significance (except in specifications 1a and 4a, which include only beta as control variable).

Table 16: Fama–Macbeth cross-sectional regressions with mean-reverting volatility

The table reports results from Fama-Macbeth cross-sectional regressions. Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll's estimator (*Roll*). Tested variables is the mean-reverting level of volatility (*m*) and its natural logarithm. R^2 reports the averaged R-squared statistics from the cross-sectional regressions. Column 'Filter' indicates the subsample used for estimating the cross-sectional regressions: 'all' in case the entire sample is used; 'UP>10' indicates specifications estimated using only securities with unadjusted price of over USD 10. Newey-West t-statistics are reported in parenthesis.

The results suggest that *m* and $\ln m$ are significant predictors of the cross-section of returns. Sign, coefficient estimates and significance levels remain stable between across alternative specifications.

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
1	all	0.36 (1.26)						0.04 (3.06)		2.76
1a	UP>10	0.24 (0.97)						0.02 (1.48)		3.37
2	all	0.41 (1.50)	0.03 (0.90)	0.88 (8.97)				0.05 (3.62)		4.07
2a	UP>10	0.29 (1.19)	0.03 (0.92)	0.6 (6.73)				0.04 (2.57)		4.75
3	all	0.31 (1.25)	0.04 (1.25)	0.88 (11.05)	0.01 (3.61)	0.02 (0.71)	-0.04 (-8.91)	0.03 (2.89)		5.78
3a	UP>10	0.28 (1.27)	0.01 (0.33)	0.63 (8.53)	0.01 (3.63)	-0.06 (-1.79)	-0.03 (-6.15)	0.04 (3.63)		6.78

#	Filter	<i>Beta</i>	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
4	all	0.26 (0.94)							0.62 (3.27)	2.78
4a	UP>10	0.16 (0.66)							0.34 (1.81)	3.40
5	all	0.30 (1.16)	0.06 (1.64)	0.90 (9.26)					0.76 (3.76)	4.14
5a	UP>10	0.19 (0.80)	0.06 (1.70)	0.63 (6.95)					0.60 (3.02)	4.81
6	all	0.23 (0.97)	0.07 (1.93)	0.9 (11.44)	0.01 (3.63)	0.01 (0.55)	-0.04 (-8.94)		0.53 (3.61)	5.84
6a	UP>10	0.19 (0.90)	0.04 (1.24)	0.65 (8.79)	0.01 (3.64)	-0.08 (-2.34)	-0.03 (-6.14)		0.67 (4.77)	6.82

Source: author's calculations

We also examine the robustness of our results in various subsamples of the original data set; these results are reported in Table 18 on p. 170. As a first test we compare the significance of idiosyncratic volatility in different market environments (volatility regimes). We calculate the mean monthly volatility of the Thomson Reuters' DataStream Total Market Return Index from daily data. Following Fink et al. (2010), we classify the volatility regimes in the market using a Markov chain with three states, where volatility in each state is assumed to be normally distributed with unknown mean and variance. The transition matrix and the parameters of each volatility state are estimated using the Baum–Welch algorithm (Baum et al., 1970), and the sequence of states of the market (the Viterbi path) is estimated using the Viterbi (1967) algorithm. The model identified three states with mean volatilities of 9.25 (the high-volatility state), 4.68 (the medium-volatility state), and 2.97 (the low-volatility state). The high-volatility state includes only 40 months, while the medium- and low-volatility states include 150 and 210 months, respectively (starting from January 1980).¹³⁰

The results from that analysis are reported as specifications 'a', 'b' and 'c' in Table 18. We find that the mean-reverting level idiosyncratic volatility is a significant predictor of the cross-section in both medium- and low-volatility episodes. The coefficient of the mean-reverting volatility is not materially different between the low-volatility and the medium-volatility state. In contrast, during episodes of market turbulence, idiosyncratic volatility made little difference for expected returns. However, it should be remarked that the number of high-volatility months (only forty) means that the results from that sub-sample should be interpreted with caution because the error of the estimates could be higher.

¹³⁰ See Chapter 3. Research Methodology and Data Sources, section "3.5. " (p. 144) for further details. The parameters of the Markov chain were provided in Table 9 on p. 147.

Table 17: Descriptive statistics by principal market, 1/1980 - 3/2013

The table summarises the descriptive statistics for the sample used in the tests, split into principal exchange where stocks are traded as assigned by Datastream.

‘Beta’ is the stock beta calculated by portfolios constructed as in Fama and French (1992). ‘ln(Cap)’ is the natural logarithm of market capitalisation calculated as number of shares times the unadjusted price. ‘ln(B/M)’ is the natural logarithm of book value of equity as available at the end of the preceding month to market price of equity (B/M), and is calculated from the price-to-book value series from Datastream (PTBV) where book values of equity are taken at a lag of six months to ensure that they are known to investors. ‘Roll’ is the bid-ask spread calculated Roll’s model; Roll (1984). \widehat{IV}^{arma} is expected volatility from ARMA(1,1) model fitted on the available series of IV_{t-1}^{daily} . ‘m’ is the mean-reverting level of volatility implied by the fitted ARMA model.

‘Mean (EW)’ and ‘Mean (VW)’ are the equally-weighted and the value-weighted values of the respective indicators. ‘St.dev.’ is the standard deviation; ‘Median’, ‘Q1’ and ‘Q3’ are the median and the first and third quartiles of the sample. ‘Skewness’ is the skewness coefficient for the sample. ‘Obs’ is the number of rows for which data is available.

The table demonstrates the differences between the Nasdaq and the non-Nasdaq sub-samples. The Nasdaq stocks are characterised with higher beta, lower liquidity (higher bid-ask spread), higher volatility, and smaller capitalisation.

Variables	Mean (EW)	Mean (VW)	St.dev.	Median	Q1	Q3	Skewness	Obs
Panel A: Nasdaq stocks								
Beta	1.30	1.19	0.30	1.28	1.06	1.50	-0.01	370,266
ln(Cap)	5.22	9.59	1.69	5.09	4.02	6.26	0.55	370,266
ln(B/M)	-0.78	-1.48	0.82	-0.71	-1.25	-0.23	-0.63	370,266
Roll	8.67	5.65	4.57	7.62	5.64	10.46	2.55	370,244
\widehat{IV}^{arma}	14.87	8.74	6.85	13.57	10.00	18.26	1.31	347,293
m	17.42	11.25	6.67	16.31	12.68	20.85	1.22	347,248

Panel B: Non-Nasdaq stocks								
Beta	1.13	0.94	0.30	1.11	0.95	1.33	0.16	493,733
$\ln(Cap)$	6.36	9.87	1.92	6.37	4.97	7.68	0.10	493,733
$\ln(B/M)$	-0.66	-1.19	0.73	-0.58	-1.03	-0.21	-0.89	493,733
Roll	5.53	3.90	3.40	4.69	3.47	6.52	3.66	493,726
\widehat{IV}^{arma}	9.18	5.98	4.88	7.97	6.01	10.90	2.37	465,627
m	10.42	6.69	5.44	9.09	6.92	12.20	2.53	465,545

Source: author's calculations

In order to examine whether idiosyncratic risk explains returns for all markets and not only of NASDAQ, which is characterised with smaller stocks with higher idiosyncratic risk, we split our sample into two sub-samples based on the exchange where the stock has primary listing¹³¹: NASDAQ and non-NASDAQ (NYSE and AMEX). The NASDAQ sub-sample includes 3,375 securities, while the non-NASDAQ sub-sample includes 3,103 securities. The results for those splits are reported as specifications ‘d’ and ‘e’ in Table 4.3. We find that the mean-reverting volatility is significant in both specifications, however the magnitude of the coefficient for non-NASDAQ securities is between one-third and one-half that for NASDAQ shares, suggesting that the differences between the two sub-samples are economically significant. In our view this finding is consistent with the prediction of the underlying theoretical model of Merton (1987), as the stocks traded at NYSE are typically larger and better-known issuers; this point is illustrated well by Table 17, which shows that Nasdaq stocks are on average more volatile and less liquid, and also characterised by higher beta. In this setting higher slope in the Fama-Macbeth regressions for Nasdaq stocks conforms with the predictions of the theoretical models that less known stocks should earn higher premium for idiosyncratic risk. According to Merton’s model, idiosyncratic risk of more widely-invested stocks should earn lower premium, which in our case is reflected in the differential between the coefficients from the Fama–MacBeth regressions on NASDAQ and non-NASDAQ sub-samples.

Another split that we perform aims to examine whether the significance of idiosyncratic volatility in explaining the cross-section holds both in periods of economic growth and during recessions. To that end we split our sample into two subsamples using the business cycle breakpoints of the National Bureau of Economic Research (NBER). These breakpoints are determined informally by the NBER’s Business Cycle Dating Committee based on a broad set of economic indicators and served as a benchmark for the state of the US economy. As customary in business cycle studies, we assumed that the peak month is classified as the last month of an expansion period, and the through month is considered as the last month of a recession period. The results are reported under specifications ‘f’ and ‘g’ in Table 18. During both expansion and recession episodes we find that idiosyncratic volatility is a significant predictor of the cross-section of stock returns, although in specification ‘3g’ the

¹³¹ Primary listing refers to the exchange where the stock was traded most actively. The allocation to a specific exchange is the one determined by Thomson-Reuters.

confidence level is somewhat lower. The differences in the point estimates for idiosyncratic risk between expansion and contraction are particularly striking – in specifications ‘3’ and ‘6’ the difference was between four- and five-fold. In general, the direction of the difference is consistent with the economic intuition that in expansion episodes investors are less risk averse and would be willing to assume more risk (in this case – idiosyncratic risk) for premium compared to contraction episodes. Nonetheless, that explanation also has caveats: firstly, the magnitude of the difference does not match the magnitude of risk aversion changes; secondly, as a baseline case one would expect that a change of risk aversion would have similar effect for the different types of risk, which is not reflected in the values of liquidity premium or beta. We consider more reasonable to attribute those results to the much smaller sample of contraction (recession) months in our test sample – only 50 months – which resulted in higher risk of the result being affected by pricing adjustments during recessions, which also overlap more with high-volatility months. Moreover, high-volatility months, for which we previously reported no correlation between idiosyncratic risk and return, are more likely to occur during recessions. Thus, 11 out of 41 high-volatility months fall during recessions, while recessions account for 56 out of 400 months, showing that a turmoil (high-volatility) month is more likely to occur during recession. Nevertheless, this difference in slopes between growth and recession is worth exploring in future studies, as it may suggest that the observed positive link between idiosyncratic risk and returns might be due to high-volatility growth companies, which earn more during a period of economic expansion, and not necessarily due to the premium on exposure to undiversified idiosyncratic risk.

Next, we investigate if the significance of idiosyncratic risk is due to previous accumulated gains or losses. The model of Merton derives the equilibrium in essentially a neo-classical framework with the added behavioural assumption that investors are constrained from holding a well-diversified portfolio. The prospect theory of investing suggests that decisions could depend on whether the position has an accumulated loss or gain relative to some reference value. Unfortunately, there is no information concerning the value used as reference by investors. Bhootra and Hur (2014), for example, employ the Capital Gains Overhang, where the reference price is measured as a weighted average of all prices in the preceding three years with weights depending on the traded volume. Benartzi and Thaler (1995) suggest that the annual period is consistent with the tax return frequency, while Malkiel and Xu (2004) point out that some institutional investors are required to disclose individual loss-making positions on a quarterly basis. Therefore, there appears to be no clear guideline as to how to evaluate the reference price used by investors. In order to investigate if

differences in accumulated gains and losses might have affect decisions, we split our sample into two sub-samples based on the accumulated gains over the six months preceding the previous month, i.e. based on our $Ret(-7, -2)$ variable, with one sub-sample including all securities with cumulative net gain over that period ($Ret(-7, -2) > 1$), while the other sub-sample including the remaining non-profit-making positions ($Ret(-7, -2) \leq 1$). These results are reported in specifications ‘h’ and ‘i’ in Table 18. The mean-reverting level of idiosyncratic volatility remains a significant predictor of the cross-section, and its magnitude does not appear to be significantly affected by whether the six-month cumulative return as measured by $Ret(-7, -2)$ is in the red or not. Nevertheless, we emphasise that these tests are intended to confirm the robustness of the mean-reverting level of volatility as a predictor of the cross-section of returns, and are not intended as a test of the prospect theory of investing as such. In particular, other measures of the reference price might yield different conclusions.

Another concern that we wish to address in this section is the possibility that idiosyncratic volatility serves as a proxy for default risk. In principle, default risk could be estimated more directly by using some combination of market and accounting data to estimate the default probabilities or at least credit risk scores. There is a plethora of models that could be useful. One of the earliest ones is Altman’s z-score (Altman, 1968), where the score Z is a linear combination of five accounting ratios, i.e.: $Z = 1.2 X_1 + 1.4 X_2 + 3.3 X_3 + 0.6 X_4 + 1.0 X_5$, where X_1 is the ratio of working capital to total assets, X_2 is the ratio of retained earnings to total assets, X_3 is the ratio of earnings before interest and taxes (EBIT) to total assets, X_4 is the ratio of market value of equity to book value of total liabilities, and X_5 is the ratio of sales to total assets. Other models build on the structural model of Merton (1974) that views the value of the firm as a stochastic process and the value of equity as a European call option on the value of the firm with exercise price equal to the value of company debt, with default occurring whenever the value process fell below the level of debt. The works of Altman and Merton inspired a number of other contributions, but due to lack of sufficient data to implement those models we shall again use filters to exclude some segments from the sample that are exposed to higher default risk. For example, Damodaran (2004) suggests (p. 256-57) that stocks that have lost substantial value over the previous year are often riskier than the remaining stocks. He points out that this is due both to the empirical regularity that low-priced stocks are more volatile, as well as to the increased leverage¹³² and financial risk

¹³² From a valuation perspective leverage should be calculated with market values of debt and

when market value of equity declines substantially. In his examination he therefore excludes stocks with price below US\$ 5, annualised volatility over 80%, beta over 1.25, and debt to market capital over 80%. For our test we implement a similar filter aimed to leave in the test sample only stocks with lower default probability: we exclude stocks with Beta over 1.25, last-month volatility (IV_{t-1}) or expected volatility (\widehat{IV}_t^{arma}) over 23.09%¹³³, and price/earnings (P/E) ratio between 12.0 and 26.0. The motivation for the use of P/E stems from its link to past earnings; when past earnings are negative, P/E is not defined, so those rows are excluded; 12 and 26 are correspondingly the first and the third quartile of the P/E data. The rationale for exclusion of low P/E securities is that the low values could signal low expected future growth or low quality of earning. The rationale for exclusion of high P/E stocks was that these may have very low earnings, which could be a signal of distress, e.g. like the securities in the dot-com bubble.¹³⁴ The cumulative application of these criteria results in a quite small subset of the original data set – only 269,685 rows remain, i.e. about 31.2% of the original data set. That subset is flagged as the “low default” subset in the table of results (specification “j”). We find that the significance and magnitude of the coefficients of the cross-sectional regression are unaffected by the restriction to the low-default-risk subsample, and in fact the coefficients are marginally higher than those of the full sample. Therefore, these tests do not provide evidence that the significance of idiosyncratic volatility is driven by its possible correlation with probabilities of default.

The impact of extreme observations should also be a concern for econometric tests. For example, Walkshäusl (2013) finds that low-volatility stocks earn abnormally high return, which the authors explain as “quality premium”, while Fu (2009) suggests that the negative

equity. Whereas market value of debt is often close to accounting value, the value of equity could drift significantly away from the accounting value, which is reflected in the price/book value ratio.

¹³³ This corresponded roughly to annual volatility of 80% since $80/\sqrt{12} \approx 23.09$.

¹³⁴ One way to think about the expected values of P/E was through the Gordon’s growth model, which proposed that the price P of a security with required rate of return (r) and growth rate (g) of dividends D_1 would be $P_0 = \frac{D_1}{r-g} = \frac{EPS_1(1-b)}{r-g}$, where b was the retention rate (the part of earnings not paid out as dividends). Noting that $g = b \times ROE$, where ROE was the return on equity, the forward price/earnings ratio took the form $P/E = \frac{P}{EPS} = \frac{1-g/ROE}{r-g}$. In that setting, a low P/E ratio could correspond to either high required rate of return (a risky business profile), or low growth perspective g . Similarly, a high P/E could be seen as result of high expected growth (low $r - g$), so that the price of the security was undepinned by growth expectations rather than by strong earnings. (Gordon and Shapiro (1956))

link documented by Ang et al. (2006) is due to return reversals of high-volatility stocks. In order to examine if our results are driven by the lowest-volatility or highest-volatility stocks, in each month we fitted the cross-sectional regression using only the volatilities between the 0.25 and 0.75 quartile in the respective month grouped by the mean-reverting level of volatility; thus, we used the middle 50% of the observation in each month, discarding the 25% with lowest volatility and the 25% with highest volatility. These results are reported as specifications “k” in Table 18. Contrary to the hypothesis that the results are driven by low- or high-volatility stocks, we find that the regression coefficients remain statistically significant and actually increase, which could suggest that the extreme volatilities add more noise to the estimates instead of driving coefficient significance.

We also explore whether the mean-reverting level of volatility could be exploited by arbitrageurs. If abnormal returns accrued only to companies that experience short episode of high volatility and expected returns quickly revert to the mean, then the transaction costs associated with exploiting differences in idiosyncratic volatilities could potentially offset the higher expected returns. For example, Li et al. (2014) reported that “in extending our tests to include holding periods beyond the first two months following portfolio formation, we found that current-month IVOL has no meaningful relationship with stock returns in periods beyond the second month. Empirically then, the excess returns of the IVOL anomaly all occur in the first month or two following portfolio formation. In short, we found that the IVOL effect is short lived, effectively requiring traders to adjust portfolio holdings at least every other month to have a reasonable chance at producing alpha. Such frequent rebalancing naturally raises questions about the impact of transaction costs and liquidity constraints” (p. 56-7). In order to examine this point, we use lagged values of mean-reverting volatilities in the cross-section of returns. If such lagged values are significant for a sufficiently long lag, then it would follow that investors could exploit the higher returns accruing to high-volatility stocks for sufficiently long periods. We used securities with unadjusted price over US\$ 10 for lags of 1, 2, 3 and 6 months. The results from that exercise are reported in Table 19 (p. 178). We find that past expected mean-reverting volatility levels remain significant predictors of the cross-section of returns for at least six months. On the other hand, the estimated coefficients for lags of six month are about one half of the coefficients of the contemporaneous coefficients.

Table 18: Fama–Macbeth cross-sectional regressions with mean-reverting volatility – robustness checks

The table reports results from Fama-Macbeth cross-sectional regressions. Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll’s estimator (*Roll*). Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. R^2 reports the averaged R-squared statistics from the cross-sectional regressions.

Column ‘Filter’ indicates the subsample used for estimating the cross-sectional regressions: ‘all’ in case the entire sample is used; ‘high vol’, ‘medium vol’ and ‘low vol’ indicate that cross-sectional regressions are estimated using only months that fall in the respective market volatility regime as classified by the Viterbi algorithms. ‘NASDAQ’ and ‘non NASDAQ’ indicate that the sample includes only securities with principal listing on NASDAQ or other (NYSE or Amex) stock exchange. ‘contraction’ and ‘expansion’ indicate that the cross-sectional regressions are estimated using months with economic contraction or expansion as classified by the National Bureau of Economic Research (NBER). ‘growing stocks’ and ‘falling stocks’ considers only stocks with cumulative growth or decline for six months based on $Ret(-7, -2)$ in order to examine possible behavioural differences depending on stock momentum. ‘low default’ indicates that the cross-sectional regressions are estimated using stocks with lower beta (below 1.25), price/earnings ratio between 12 and 26, and lower values of last-month and current-month ARMA forecasts below 23%.

The results suggest that *m* and $\ln m$ are significant predictors of the cross-section of returns. Sign, coefficient estimates and significance levels remain stable across alternative specifications. Premium on NASDAQ stocks, however, tends to be notably higher. However, premium during contraction episodes appears to be significantly higher compared to periods of expansion – a difference that cannot be explained solely as changes in risk tolerance.

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
1	all	0.36						0.04		2.76
		(1.26)						(3.06)		
1a	high vol	-1.70						-0.03		5.12
		(-1.46)						(-0.49)		
1b	medium vol	1.45						0.06		2.99
		(2.81)						(2.64)		
1c	low vol	0.01						0.04		2.15
		(0.05)						(2.77)		
1d	NASDAQ	0.35						0.06		1.85
		(0.95)						(3.29)		

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
1e	non-NASDAQ	0.37 (1.34)						0.04 (3.29)		2.86
1f	contraction	0.44 (0.47)						0.08 (2.02)		3.41
1g	expansion	0.35 (1.25)						0.04 (2.55)		2.66
1h	growing stocks	0.48 (1.62)						0.06 (4.31)		3.07
1i	falling stocks	0.14 (0.46)						0.04 (2.38)		2.68
1j	low default	0.27 (1.01)						0.04 (2.47)		2.18
1k	low default	0.28 (0.98)						0.06 (2.48)		1.66
2	all	0.41 (1.50)	0.03 (0.90)	0.88 (8.97)				0.05 (3.62)		4.07
2a	high vol	-1.84 (-1.69)	0.08 (0.71)	1.58 (3.54)				-0.02 (-0.30)		7.36
2b	medium vol	1.44 (2.96)	0.05 (0.90)	0.82 (4.48)				0.06 (2.79)		4.52

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
2c	low vol	0.13	0.00	0.79				0.05		3.14
		(0.47)	(0.08)	(8.43)				(3.21)		
2d	NASDAQ	0.59	0.20	1.29				0.09		3.32
		(1.72)	(2.90)	(10.19)				(4.28)		
2e	non-NASDAQ	0.27	0.00	0.72				0.04		4.15
		(1.05)	(-0.12)	(7.96)				(3.29)		
2f	contraction	0.29	0.02	0.97				0.08		4.88
		(0.33)	(0.24)	(2.63)				(2.09)		
2g	expansion	0.43	0.03	0.87				0.04		3.95
		(1.57)	(0.86)	(8.91)				(3.01)		
2h	growing stocks	0.54	-0.04	0.71				0.05		4.39
		(1.86)	(-1.04)	(7.30)				(4.15)		
2i	falling stocks	0.12	0.06	1.13				0.05		3.98
		(0.42)	(1.63)	(11.94)				(3.06)		
2j	low default	0.23	0.06	0.36				0.05		3.50
		(0.82)	(1.64)	(4.07)				(3.01)		
2k	middle 50%	0.24	0.10	0.90				0.09		3.33
		(0.93)	(2.18)	(8.35)				(3.19)		

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
3	all	0.31 (1.25)	0.04 (1.25)	0.88 (11.05)	0.01 (3.61)	0.02 (0.71)	-0.04 (-8.91)	0.03 (2.89)		5.78
3a	high vol	-1.72 (-1.77)	0.06 (0.52)	1.16 (3.53)	-0.01 (-0.94)	-0.13 (-1.17)	-0.05 (-2.99)	0.01 (0.28)		11.28
3b	medium vol	1.15 (2.71)	0.07 (1.11)	0.87 (5.85)	0.01 (3.09)	0.07 (1.52)	-0.04 (-5.41)	0.03 (1.78)		6.31
3c	low vol	0.12 (0.48)	0.02 (0.52)	0.84 (9.54)	0.01 (4.77)	0.02 (0.50)	-0.03 (-7.67)	0.03 (2.39)		4.37
3d	NASDAQ	0.52 (1.69)	0.20 (3.01)	1.25 (10.21)	0.02 (4.66)	0.00 (-0.11)	-0.05 (-7.58)	0.07 (2.82)		5.94
3e	non-NASDAQ	0.12 (0.53)	0.00 (0.07)	0.73 (9.94)	0.01 (2.55)	0.03 (1.04)	-0.03 (-6.93)	0.02 (2.20)		6.26
3f	contraction	0.17 (0.20)	0.04 (0.39)	0.93 (4.31)	0.00 (0.14)	-0.14 (-1.48)	-0.05 (-3.62)	0.11 (3.10)		7.4
3g	expansion	0.33 (1.34)	0.04 (1.16)	0.88 (10.13)	0.01 (5.36)	0.04 (1.52)	-0.03 (-8.29)	0.02 (1.84)		5.55
3h	growing stocks	0.38 (1.47)	-0.03 (-0.94)	0.77 (8.92)	0.01 (6.09)	-0.04 (-1.17)	-0.02 (-3.17)	0.04 (3.55)		5.94
3i	falling stocks	-0.01 (-0.02)	0.10 (2.49)	1.00 (12.18)	0.00 (0.96)	0.06 (1.65)	-0.05 (-10.97)	0.04 (2.30)		6.04

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
3j	low default	0.16 (0.60)	0.05 (1.20)	0.36 (4.38)	0.00 (1.69)	-0.02 (-0.43)	-0.05 (-9.46)	0.05 (3.02)		6.01
3k	middle 50%	0.11 (0.49)	0.11 (2.46)	0.92 (10.44)	0.01 (3.02)	-0.02 (-0.62)	-0.04 (-8.06)	0.09 (3.77)		5.31
4	all	0.26 (0.94)							0.62 (3.27)	2.78
4a	high vol	-1.66 (-1.56)							-0.41 (-0.41)	5.16
4b	medium vol	1.26 (2.58)							0.93 (2.99)	3.04
4c	low vol	-0.06 (-0.23)							0.60 (3.15)	2.14
4d	NASDAQ	0.32 (0.88)							0.86 (3.35)	1.73
4e	non-NASDAQ	0.25 (0.94)							0.59 (3.75)	2.91
4f	contraction	0.25 (0.28)							1.07 (2.14)	3.43
4g	expansion	0.26 (0.97)							0.55 (2.72)	2.68

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
4h	growing stocks	0.30 (1.08)							0.88 (4.61)	3.09
4i	falling stocks	0.11 (0.36)							0.51 (2.43)	2.65
4j	low default	0.19 (0.71)							0.42 (2.65)	2.25
4k	middle 50%	0.28 (0.98)							0.73 (2.48)	1.65
5	all	0.30 (1.16)	0.06 (1.64)	0.90 (9.26)					0.76 (3.76)	4.14
5a	high vol	-1.8 (-1.83)	0.06 (0.50)	1.57 (3.42)					-0.31 (-0.30)	7.58
5b	medium vol	1.23 (2.71)	0.11 (1.65)	0.86 (4.86)					1.16 (3.02)	4.62
5c	low vol	0.06 (0.21)	0.02 (0.54)	0.80 (8.71)					0.68 (3.58)	3.16
5d	NASDAQ	0.51 (1.57)	0.22 (3.05)	1.30 (10.33)					1.40 (4.44)	3.20
5e	non-NASDAQ	0.16 (0.65)	0.02 (0.74)	0.74 (8.18)					0.60 (3.83)	4.23

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
5f	contraction	0.11 (0.14)	0.08 (0.71)	1.01 (2.72)					1.21 (2.16)	4.96
5g	expansion	0.33 (1.28)	0.05 (1.42)	0.88 (9.25)					0.69 (3.10)	4.02
5h	growing stocks	0.37 (1.41)	0.00 (0.00)	0.74 (7.86)					0.89 (4.34)	4.46
5i	falling stocks	0.09 (0.33)	0.07 (1.60)	1.13 (11.83)					0.66 (2.87)	4.01
5j	low default	0.12 (0.43)	0.08 (1.93)	0.37 (4.19)					0.64 (3.27)	3.54
5k	middle 50%	0.25 (0.94)	0.10 (2.16)	0.90 (8.32)					1.01 (3.03)	3.32
6	all	0.23 (0.97)	0.07 (1.93)	0.90 (11.44)	0.01 (3.63)	0.01 (0.55)	-0.04 (-8.94)		0.53 (3.61)	5.84
6a	high vol	-1.71 (-1.86)	0.05 (0.37)	1.15 (3.42)	-0.01 (-0.92)	-0.13 (-1.16)	-0.05 (-3.00)		0.09 (0.16)	11.43
6b	medium vol	1 (2.49)	0.12 (1.75)	0.91 (6.32)	0.01 (3.12)	0.05 (1.22)	-0.04 (-5.45)		0.74 (2.74)	6.38
6c	low vol	0.07 (0.28)	0.04 (1.00)	0.85 (9.85)	0.01 (4.81)	0.02 (0.55)	-0.03 (-7.71)		0.48 (2.77)	4.4

#	Filter	<i>Beta</i>	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
6d	NASDAQ	0.48 (1.64)	0.22 (3.13)	1.27 (10.33)	0.02 (4.67)	0.00 (0.07)	-0.05 (-7.45)		1.04 (3.45)	5.79
6e	non-NASDAQ	0.05 (0.21)	0.03 (0.83)	0.75 (10.19)	0.01 (2.56)	0.02 (0.79)	-0.03 (-6.97)		0.43 (3.07)	6.32
6f	contraction	-0.02 (-0.02)	0.10 (0.84)	0.97 (4.46)	0.00 (0.14)	-0.14 (-1.59)	-0.05 (-3.64)		1.55 (3.01)	7.45
6g	expansion	0.26 (1.13)	0.06 (1.66)	0.89 (10.54)	0.01 (5.41)	0.04 (1.39)	-0.03 (-8.32)		0.39 (2.57)	5.6
6h	growing stocks	0.26 (1.07)	0.00 (0.03)	0.80 (9.44)	0.01 (6.10)	-0.04 (-1.42)	-0.02 (-3.27)		0.69 (4.54)	5.98
6i	falling stocks	-0.02 (-0.08)	0.11 (2.50)	1.00 (12.24)	0.00 (0.97)	0.06 (1.64)	-0.05 (-11.07)		0.43 (2.13)	6.05
6j	low default	0.07 (0.25)	0.06 (1.48)	0.37 (4.51)	0.01 (1.79)	-0.02 (-0.62)	-0.05 (-9.55)		0.63 (3.47)	5.99
6k	middle 50%	0.11 (0.50)	0.10 (2.44)	0.92 (10.41)	0.01 (3.02)	-0.02 (-0.60)	-0.04 (-8.05)		0.98 (3.69)	5.30

Source: author's calculations

Table 19: Fama-Macbeth cross-sectional regressions with mean-reverting volatility – return persistence

The table reports results from Fama-Macbeth cross-sectional regressions using lagged values of the mean-reverting level of volatility (m_{-l}) and its natural logarithm. If the mean-reverting level allows construction of tradable strategies, we would like the predictive performance of the variable to remain stable in order to reduce the costs of portfolio rebalancing. The lag used in the test is indicated in column ‘Lag (l)’.

Control variables are beta with market ($Beta$), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll’s estimator ($Roll$). ‘ R^2 ’ reports the averaged R-squared statistics from the cross-sectional regressions.

The results suggest that m and $\ln m$ are significant predictors of the cross-section of returns for lags up to 6 months. The regression coefficient remains statistically significant, but its value declines with the increase of the lag.

#	Lag (l)	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	m_{-l}	$\ln m_{-l}$	R^2
1	$l = 1$	0.22 (0.87)						0.03 (1.78)		3.43
	$l = 2$	0.28 (1.11)						0.02 (1.16)		3.48
	$l = 3$	0.29 (1.16)						0.02 (0.92)		3.52
	$l = 6$	0.28 (1.15)						0.01 (0.72)		3.59
2	$l = 1$	0.27 (1.10)	0.04 (1.17)	0.62 (6.88)				0.05 (2.95)		4.81
	$l = 2$	0.32 (1.31)	0.03 (0.86)	0.60 (6.79)				0.04 (2.18)		4.86
	$l = 3$	0.34 (1.37)	0.02 (0.73)	0.60 (6.70)				0.03 (1.80)		4.92

#	Lag (l)	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	m_{-l}	$\ln m_{-l}$	R^2
	$l = 6$	0.33 (1.36)	0.02 (0.73)	0.59 (6.53)				0.03 (1.56)		5.01
3	$l = 1$	0.26 (1.22)	0.02 (0.57)	0.64 (8.68)	0.01 (3.61)	-0.07 (-1.97)	-0.03 (-6.21)	0.05 (4.11)		6.84
	$l = 2$	0.31 (1.40)	0.01 (0.31)	0.63 (8.62)	0.01 (3.69)	-0.05 (-1.58)	-0.03 (-6.33)	0.03 (2.74)		6.91
	$l = 3$	0.31 (1.42)	0.01 (0.22)	0.63 (8.49)	0.01 (3.80)	-0.05 (-1.36)	-0.03 (-6.50)	0.03 (2.18)		6.99
	$l = 6$	0.30 (1.40)	0.01 (0.21)	0.62 (8.25)	0.01 (3.98)	-0.05 (-1.50)	-0.03 (-6.61)	0.02 (1.94)		7.11
4	$l = 1$	0.13 (0.56)							0.39 (2.00)	3.46
	$l = 2$	0.20 (0.85)							0.28 (1.44)	3.50
	$l = 3$	0.22 (0.94)							0.23 (1.17)	3.54
	$l = 6$	0.24 (1.01)							0.18 (0.87)	3.62

#	Lag (l)	$Beta$	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	$Roll$	Ret_{t-1}	m_{-l}	$\ln m_{-l}$	R^2
5	$l = 1$	0.16 (0.68)	0.07 (1.96)	0.64 (7.11)					0.68 (3.27)	4.88
	$l = 2$	0.23 (0.98)	0.05 (1.53)	0.62 (6.96)					0.54 (2.55)	4.93
	$l = 3$	0.25 (1.08)	0.05 (1.34)	0.62 (6.85)					0.47 (2.17)	4.98
	$l = 6$	0.27 (1.17)	0.04 (1.12)	0.60 (6.56)					0.39 (1.75)	5.07
6	$l = 1$	0.17 (0.81)	0.05 (1.52)	0.67 (8.96)	0.01 (3.63)	-0.09 (-2.52)	-0.03 (-6.19)		0.76 (5.05)	6.88
	$l = 2$	0.23 (1.10)	0.04 (1.08)	0.65 (8.83)	0.01 (3.70)	-0.07 (-2.07)	-0.03 (-6.32)		0.56 (3.69)	6.95
	$l = 3$	0.24 (1.16)	0.03 (0.88)	0.65 (8.68)	0.01 (3.80)	-0.06 (-1.77)	-0.03 (-6.48)		0.47 (3.00)	7.02
	$l = 6$	0.26 (1.24)	0.02 (0.68)	0.63 (8.32)	0.01 (4.00)	-0.06 (-1.79)	-0.03 (-6.60)		0.38 (2.42)	7.15

Source: author's calculations

4.5. Further tests of robustness

4.5.1. Was there an omitted factor?

The evidence presented thus far suggests that the mean-reverting level of idiosyncratic volatility is a significant predictor of the cross-section of returns. Nevertheless, as discussed previously, one measures idiosyncratic risk with respect to some specific factor model – in our case, the one of Fama–French–Carhart. It may happen that idiosyncratic volatility serves only as a proxy for the loading on some other, omitted factor. For example, we found that the premium on idiosyncratic risk was lower during an economic downturn. Then we may hypothesise that our model lacks some economic factor, and the inclusion of the loading on that factor in the cross-sectional regressions could make the idiosyncratic risk insignificant. On the other hand, if the Merton model is correct, investors should dislike idiosyncratic risk purely because of under-diversification and not because high-volatility stocks have greater exposure to some systematic factor.

In this section we address that problem by estimating the principal omitted factor and adding its loading as an explanatory variable in the cross-sectional regressions. If the principal factor is the reason for the significance of the idiosyncratic risk variable, then we expect that the addition of the new loading will make the slope of idiosyncratic risk insignificant. Firstly, we use statistical factor analysis¹³⁵ in order to estimate the returns on the omitted factor(s). We then use the loadings of the individual securities on that factor (or, possible, on a number of factors) in our cross-sectional regressions.

More specifically, we use the heteroscedastic factor analysis of Jones (2001) in order to extract a set of K common factors underlying observed idiosyncratic returns ($IRet_{i,t}$). The common factors are extracted from the full data set, comprising all available time series of idiosyncratic innovations for the studied period. The number of factors can be selected using the PC_{p1} and PC_{p2} criteria discussed by Bai and Ng (2002). Then, factor loadings are estimated as the coefficients of ordinary least squares (OLS) regressions of monthly returns on each stock on the set of factors identified at the preceding stage of the analysis.

Let R^n denote the $n \times T$ matrix of observed idiosyncratic returns ($IRet_{i,t}$), H - the

¹³⁵ For a review refer to Ch.4 in Connor et al. (2010)

matrix of (unobservable) factor realisations, B^n - the matrix of factor loading, and E^n - residual returns. These residual returns could be viewed as the remaining idiosyncratic innovations after accounting for the omitted factors of idiosyncratic returns. Then the model of idiosyncratic returns is assumed to be in the form:

$$R^n = B^n H + E^n.$$

Let $F \equiv M^{-1/2}H$ stand for the matrix of rotated factor realisations, introduced to simplify notation, and let D denote the diagonal matrix of average idiosyncratic variances. Jones (2001) proves that the average variance $(1/n)R^{n'}R^n$ converges to $F'F + D$ and could be estimated using Jöreskog's procedure, i.e.:

1. Compute $C = (1/n)R^{n'}R^n$;
2. Guess an initial D , e.g. $D = 0.5C$;
3. Obtain the K largest eigenvalues of $D^{-1/2}CD^{-1/2}$ and create a diagonal matrix Λ having the largest eigenvalues along its main diagonal; then create a matrix V of their corresponding eigenvectors;
4. Estimate the factor matrix as $F = D^{1/2}V(\Lambda - I)^{1/2}$;
5. Compute a new estimate of $D = C - F'F$ and return to 3 until the algorithm converged;
6. Estimate factor loading using OLS regression of observed excess returns on estimated factors and obtain residual idiosyncratic errors ϵ .

An important issue in factor analysis is the choice of an appropriate number of factors. Remembering that we are analysing residuals (idiosyncratic returns) obtained from a factor model that already has three factors, we perform this test assuming only one omitted common factor. That factor would be the one with the highest contribution in explaining the common pattern of idiosyncratic residuals. The methodology could be readily expanded further with the inclusion of more than one factor.¹³⁶

Summary statistics for the recovered factor and the loading of idiosyncratic returns on the extracted factor (f_{hfa}) are given in Table 20. Factor loadings (β_{hfa}) are estimated using rolling regressions with monthly idiosyncratic returns ($IRet_{i,t}$) from the last 24 to 60 months preceding the current month, as available (i.e. from $(t - 1)$ until $(t - k)$, $k = 24 \dots 60$).

¹³⁶ There is no commonly accepted approach to selecting the number of factors. In general, such approaches aim to measure whether an additional factor has additional explanatory power (e.g. Kandel and Stambaugh (1989), Connor and Korajczyk, and Bai and Ng). For further details see Kandel and Stambaugh (1989); Connor and Korajczyk (1993); Bai and Ng (2002)

Table 20: Summary statistics of the first common factor of idiosyncratic returns and loadings on that factor, 07/1982-03/2013

The table reports summary statistics of the latent (statistical) factor (f_{hfa}) estimated using the heteroscedastic factor analysis of Jones (2001) and the loading (β_{hfa}) on the statistical factor. ‘Mean (EW)’ and ‘Mean (VW)’ are equally-weighted and value-weighted averages, ‘Std.dev’ stands for the standard deviation of the sample., and ‘Median’, ‘Q1’ and ‘Q3’ stand for the second, first and third quartiles of the sample. ‘Skew’ is the skewness coefficient of the sample, and ‘Count’ is the number of records in the sample.

	Mean (EW)	Mean (VW)	Std.dev	Median	Q1	Q3	Skew	Count
f_{hfa}	-0.16	-	5.05	-0.06	-3.25	3.14	0.00	393
β_{hfa}	0.05	0.03	0.48	-0.01	-0.23	0.26	0.79	713,860

Source: author’s calculations

Table 21 reports the results from Fama–Macbeth cross-sectional regressions involving the loading (β_{hfa}) on the statistical factor (f_{hfa}) explaining asset returns. We find that although the coefficient of the loading is always negative, it is not significant at conventional levels in any of the specifications. On the other hand, the coefficients for both the mean-reverting level of volatility (m) and its natural logarithm ($\ln m$) are significant and positive, consistent with the results from the previous section. The results support the conclusion that the mean-reverting level of volatility is a significant predictor of the cross-section of returns and its significance is not attributable to correlation with the loading on some unknown factor that affects idiosyncratic returns but was omitted from the Fama–French–Carhart specification.

Table 21: Fama–Macbeth cross-sectional regressions with loading on the principal factor affecting idiosyncratic returns, 07/1982 – 03/2013

The table reports the results from Fama-Macbeth cross-sectional regressions involving the loading on the statistical factor (β_{hfa}). Control variables are beta with market (*Beta*), natural logarithms of market capitalisation (*ln(Cap)*) and book-to-market value (*ln(B/M)*), momentum of returns measured as cumulative return from (t-7) until (t-1) (*Ret(-7, -2)*), return in the previous month (*Ret_{t-1}*), liquidity measured in terms of Roll’s estimator (*Roll*). *R*² reports the averaged R-squared statistics from the cross-sectional regressions. *m* and *ln m* denote the mean-reverting level of volatility and its natural logarithm. The table shows that the statistical factor is insignificant in explaining the cross-section beyond the contribution of the volatility forecast, and albeit its value remains negative and reasonably stable, it is found to be statistically insignificant.

#	<i>Beta</i>	<i>ln(Cap)</i>	<i>ln(B/M)</i>	<i>Ret(-7, -2)</i>	<i>Roll</i>	<i>Ret_{t-1}</i>	<i>m</i>	<i>ln m</i>	β_{hfa}	<i>R</i> ²
1	0.91 (2.55)								-0.06 (-0.37)	2.73
2	0.82 (2.53)	-0.04 (-1.14)	0.87 (8.21)						-0.09 (-0.53)	4.24
3	0.54 (1.93)	0.03 (0.92)	0.88 (10.65)	0.01 (2.87)	0.06 (2.60)	-0.04 (-9.58)			-0.18 (-1.14)	6.01
4	0.47 (1.79)	0.06 (1.76)	0.90 (11.41)	0.01 (2.59)	0.03 (1.29)	-0.04 (-9.66)	0.02 (2.22)		-0.17 (-1.04)	6.42
5	0.39 (1.55)	0.08 (2.33)	0.91 (11.69)	0.01 (2.59)	0.03 (1.11)	-0.04 (-9.67)		0.44 (3.12)	-0.18 (-1.12)	6.47

Source: author’s calculations

4.5.2. Evidence from daily data

We have previously pointed out that the Capital Asset Pricing Model (CAPM) and its extension – Merton’s model – are not linked to any specific time frequency. Consequently, results that hold on daily data should also hold at lower frequencies, e.g. weekly, monthly, quarterly, or annually. The opposite is also true: if the mean-reverting level of volatility is explaining the cross-section of returns at monthly frequency, then a similar effect should be observed at daily frequency.

For the purposes of the test, the change of frequency, however, is not a trivial task. Models that are suited for monthly frequency by accommodating the shorter time-series and the higher noise in the observed proxies of the underlying, unobservable idiosyncratic volatility, could be unsuitable for higher frequencies. For example, the previously referenced study of Hansen and Lunde (2005) documents the strong empirical performance of the simple GARCH(1,1) for one-day forecasts. On the other hand, the simple model also has certain drawbacks, which invited the development of a host of GARCH extensions that accommodate phenomena observed in various financial time series.¹³⁷ For example, the observation that volatility reacts differently to positive and negative shocks invited the development of the Exponential GARCH model¹³⁸. The idea of different volatility regimes led to the development of the Regime-Switching ARCH model, where parameter values depend on the volatility regime.¹³⁹ The observation of slow decay of the auto-covariance function of empirical volatility processes prompted the development of the Fractionally Integrated GARCH model.¹⁴⁰

In our study we emphasise the role of the mean-reverting level of idiosyncratic volatility, to which current volatility is expected to converge gradually over time. We also found that short-horizon forecasts (one-month-ahead) were insignificant predictors of the cross-section of returns, which in the present context also suggests that we should aim to employ a model that has long memory in volatility and which will not be unduly affected by short-term volatility outbursts. Such considerations could suggest the use of the FIGARCH or RS-(G)ARCH models, both of which offer interesting approaches for capturing the

¹³⁷ see Pagan and Schwert (1990) for a comparison of competing volatility forecasting techniques

¹³⁸ see Nelson (1991)

¹³⁹ see Hamilton and Susmel (1994)

¹⁴⁰ Baillie et al. (1996)

mean-reverting level of volatility. Structurally, however, we think that the Component GARCH (CGARCH) model of Lee and Engle (1999) offers a more direct and transparent route to capturing the mean-reverting volatility while allowing the mean-reverting level to evolve over time. When we used monthly data, such specification was not feasible due to scarcity of data, but at the daily frequency it offers a more direct implementation of our approach.

Lee and Engle propose to model volatility as the sum of two components: a permanent one and a transitory one. The permanent component has persistence close to unity, whereas the transient component has an expected value of zero and decays more quickly. More specifically, we employ Component GARCH(1,1), which has the following specification:

$$\begin{aligned}\sigma_t^2 &= q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}), \\ q_t &= \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2),\end{aligned}$$

where q_t is the permanent component of volatility, while $(\sigma_t^2 - q_t)$ is the transitory component of volatility. The n -step expected values from the CGARCH(1,1) are found to be as follows:

$$\mathbb{E}_{t-1}(\sigma_{t+n}^2) = \mathbb{E}_{t-1}(q_{t+n}) + (\alpha + \beta)^n(\sigma_t^2 - q_t),$$

and

$$\mathbb{E}_{t-1}(q_{t+n}) = \frac{1 - \rho^n}{1 - \rho} \omega + \rho^n q_t,$$

and when $n \rightarrow \infty$, the unconditional expected values became $\mathbb{E}_{t-1}(\sigma_{t+n}^2) = \mathbb{E}_{t-1}(q_{t+n}) = \frac{\omega}{1 - \rho}$.

These formulae allow us to test our results from the previous section. Firstly, we can calculate the forecasted volatility for one month ahead, which we can take to be 21 days, so that the one-month forecast would be $\mathbb{E}_{t-1}(\sigma_{t+21}^2)$, which would be an equivalent of the volatility forecasts from the previous section. We can also calculate the unconditional expectation of the permanent component, towards which volatility is expected to revert as $n \rightarrow \infty$, i.e. $\mathbb{E}_{t-1}(q_{t+\infty}) = \omega/(1 - \rho)$. We could then use these two volatility forecasts to analyse the cross-section of monthly¹⁴¹ returns. Based on our results thus far we expect that $\mathbb{E}_{t-1}(\sigma_{t+21}^2)$ would not be a significant predictor of the cross-section of returns, while $\mathbb{E}_{t-1}(q_{t+\infty})$ would be significant.

¹⁴¹ Of course nothing prevents the use of the same approach to test its performance in explaining other frequencies, e.g. daily or weekly returns. At any rate, the use of daily returns should also take into account the possible impact of market microstructure effects.

More specifically, we proceed as follows. We select a random sample of 2100 securities from the full sample of available securities.¹⁴² For each security before the start of each month we calculate daily idiosyncratic returns from the Fama–French–Carhart model using the last five years of data, as available, but not less than 250 returns.¹⁴³ We estimate a CGARCH(1,1) model using the available history from the rolling window, and forecast the expected volatility at the end of the month ($\mathbb{E}_{t-1}(\sigma_{t+21}^2)$), as well as the unconditional expected value of the mean level ($\mathbb{E}_{t-1}(q_{t+\infty})$).¹⁴⁴

Table 22: Summary statistics of expected volatilities from daily data

The table reports summary statistics for estimates of next-month volatility ($\mathbb{E}_{t-1}(\sigma_{t+21}^2)$) and of unconditional volatility ($\mathbb{E}_{t-1}(q_{t+\infty})$) estimated directly from daily returns using Component GARCH(1,1). Estimates are calculated using expanding window design for a random sample of 2100 securities. Volatilities are scaled to monthly frequency. The table reports values for equally-weighted and value-weighted averages (‘Mean (EW)’ and ‘Mean (VW)’), standard deviation (‘St.dev.’), median (‘Median’), first and third quartiles (‘Q1’ and ‘Q3’), skewness coefficient (‘Skewness’) and number of observations (‘Obs’).

Variables	Mean (EW)	Mean (VW)	St.dev.	Median	Q1	Q3	Skewness	Obs
$\mathbb{E}_{t-1}(\sigma_{t+21}^2)$	14.41	8.11	8.83	12.06	8.32	17.99	2.13	278,301
$\ln \mathbb{E}_{t-1}(\sigma_{t+21}^2)$	2.52	1.99	0.54	2.49	2.12	2.89	0.23	278,301
$\mathbb{E}_{t-1}(q_{t+\infty})$	18.41	9.68	16.66	13.73	8.93	22.10	4.09	278,301
$\ln \mathbb{E}_{t-1}(q_{t+\infty})$	2.66	2.11	0.68	2.62	2.19	3.10	0.29	278,301

Source: author’s calculations

Our results are presented in Table 23. Consistent with the results from the preceding sections, we find that expected volatility for the next month, $\mathbb{E}_{t-1}(\sigma_{t+21}^2)$, is not a significant predictor of the cross-section of returns. On the other hand, the mean-reverting level $\mathbb{E}_{t-1}(q_{t+\infty})$ is a significant predictor. The slopes of $\mathbb{E}_{t-1}(q_{t+\infty})$ and $\ln \mathbb{E}_{t-1}(q_{t+\infty})$ are somewhat lower than those estimated from monthly ARMA(1,1), which were reported

¹⁴² The use of sampling is motivated by the high computational burden of calculating expected volatilities while preventing look-ahead bias. A small sample results in a lower number securities available in each month, and hence high errors of the cross-sectional coefficients, which in turn translates into higher likelihood of failing to reject the null hypothesis that the coefficients of the cross-sectional regressions are insignificant.

¹⁴³ We excluded daily returns below -25% or above 300% as possible data errors, as suggested in Ince and Porter (2006)

¹⁴⁴ For easier comparison with the coefficients from the preceding section, the daily volatilities are scaled to monthly frequency by multiplying by a constant factor of $\sqrt{21}$.

previously in this thesis. Part of the reason could be the scaling from daily to monthly frequency, which is evident in the higher average expected idiosyncratic volatilities reported in Table 6 (p. 117), compared with the summary statistics of the series obtained directly from daily data, reported in Table 22. Nevertheless, the magnitude of the difference is such that it could not be explained solely by scaling. One plausible explanation could be the lower persistence of daily volatilities compared to monthly volatilities, which may somewhat dilute the predictive performance of the mean-reverting level estimated from daily data. On the other hand, the predictive performance of the mean-reverting volatility remains robust in the alternative model specifications, which confirmed its relevance in explaining the cross-section of returns.

Table 23: Cross-sectional Fama–Macbeth regressions with volatilities calculated from daily data using Component GARCH(1,1)

The table reports results from Fama-Macbeth cross-sectional regressions using volatility forecasts obtained from Component GARCH(1,1) model fitted on expanding window of daily returns. $\mathbb{E}_{t-1}(\sigma_{t+21}^2)$ is the estimate of next-month volatility, and $\mathbb{E}_{t-1}(q_{t+\infty})$ denotes the corresponding unconditional volatility for the model. Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(Cap)$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($Ret(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll’s estimator (*Roll*). R^2 reports the averaged R-squared statistics from the cross-sectional regressions. m and $\ln m$ denote the mean-reverting level of volatility and its natural logarithm. The results are consistent with those obtained from monthly data. One-month forecasts are not statistically significant once momentum ($Ret(-7, -2)$), liquidity (*Roll*) and return reversals (Ret_{t-1}) are added to the model. On the other hand, unconditional volatility remains a significant predictor of the cross-section. However, the estimate of the coefficient for $\mathbb{E}_{t-1}(q_{t+\infty})$ is markedly below the corresponding value when forecasts from ARMA(1,1) are used, highlighting the room for improvement of the specification.

#	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	$\mathbb{E}_{t-1}(\sigma_{t+21}^2)$	$\ln \mathbb{E}_{t-1}(\sigma_{t+21}^2)$	$\mathbb{E}_{t-1}(q_{t+\infty})$	$\ln \mathbb{E}_{t-1}(q_{t+\infty})$	R^2
1	0.47 (1.63)						0.03 (3.02)				2.83
2	0.58 (2.07)	-0.01 (-0.21)	0.90 (7.67)				0.03 (2.24)				4.35
3	0.49 (0.25)	0.01 (0.04)	0.92 (0.08)	0.01 (0.00)	0.07 (0.03)	-0.03 (-0.00)	0.00 (0.01)				6.11
4	0.34 (1.27)							0.64 (3.04)			2.76
5	0.48 (1.88)	0.00 (0.05)	0.90 (7.67)					0.54 (2.36)			4.34
6	0.47 (1.91)	0.01 (0.34)	0.91 (9.04)	0.01 (4.06)	0.06 (2.28)	-0.03 (-7.90)		0.11 (0.56)			6.15
7	0.60 (1.86)								0.02 (3.58)		2.37

#	<i>Beta</i>	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	<i>Roll</i>	Ret_{t-1}	$\mathbb{E}_{t-1}(\sigma_{t+21}^2)$	$\ln \mathbb{E}_{t-1}(\sigma_{t+21}^2)$	$\mathbb{E}_{t-1}(q_{t+\infty})$	$\ln \mathbb{E}_{t-1}(q_{t+\infty})$	R^2
8	0.60 (1.96)	-0.02 (-0.49)	0.92 (7.37)						0.02 (4.49)		3.96
9	0.42 (1.58)	0.03 (0.84)	0.94 (9.34)	0.01 (3.86)	0.05 (1.77)	-0.03 (-7.77)			0.01 (3.44)		5.99
10	0.44 (1.47)									0.47 (3.77)	2.43
11	0.47 (1.64)	0.01 (0.27)	0.94 (7.73)							0.52 (4.48)	3.99
12	0.36 (1.38)	0.05 (1.21)	0.96 (9.58)	0.01 (3.87)	0.05 (1.84)	-0.04 (-7.83)				0.30 (3.70)	5.99

Source: author's calculations

4.5.3. Portfolios as assets

The preceding section utilised the Fama–MacBeth methodology using individual securities as assets. The rationale for that choice stems from the superior efficiency compared to portfolios¹⁴⁵ and is consistent with the approaches of Fama and French (1992) and Fu (2009). From a practical perspective, however, it is also useful to examine the role of idiosyncratic volatility using suitably constructed portfolios. This allows us to drill down the results and identify situations where the mean-reverting volatility is likely to be useful and situations where it may not be appropriate.

We form double-sorted portfolios based on capitalisation and the expected mean-reverting level of volatility. We first form ten decile portfolios based on market capitalisation, and then split each of these into five quintile portfolios based on the mean-reverting level of volatility. For each portfolio we calculate the simple (equal-weighted) average return for the next one month, after which the double-sorting procedure is repeated. In order to prevent the numerous small-capitalisation stocks listed at NASDAQ from asserting undue influence on our estimates we employ breakpoints calculated only from NYSE securities. The choice of the number of portfolios is driven by pragmatic considerations: the significant correlation between the explanatory variables (beta, size, liquidity, and idiosyncratic volatility) necessitates a higher number of portfolios, so that using only twenty-five portfolios appears unjustified. On the other hand, using decile NYSE breakpoints for the mean-reverting level of volatility results into too few (fewer than ten) securities in some portfolios in the earlier months of the sample. Therefore, we opt for using fifty portfolios for this test.

¹⁴⁵ Ang et al. (2010)

Table 24: Fama–Macbeth regressions with quintile portfolios as assets

The table reports results from Fama-Macbeth cross-sectional regressions calculated on a total of 50 double-sorted portfolios formed each month by capitalisation (ten decile portfolios) and idiosyncratic risk (five quintile portfolios) using breakpoints for NYSE stocks only. For each portfolio we calculate the simple (equal-weighted) average return for the next month, after which the double-sorting procedure is repeated.

Control variables are beta with market (*Beta*), natural logarithms of market capitalisation (*ln(Cap)*) and book-to-market value (*ln(B/M)*), momentum of returns measured as cumulative return from (t-7) until (t-1) (*Ret(-7, -2)*), liquidity measured in terms of Roll’s estimator (*Roll*). Control variables are averaged by portfolios. Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. *R*² reports the averaged R-squared statistics from the cross-sectional regressions.

Column ‘Filter’ indicates the subsample used for estimating the cross-sectional regressions: ‘all’ in case the entire sample is used; ‘UP>10’ indicates that cross-sectional regressions are estimated using only stocks with unadjusted price above USD 10; ‘NASDAQ’ and ‘non NASDAQ’ indicate that the sample includes only securities with principal listing on NASDAQ or other (NYSE or Amex) stock exchange.

The results suggest that *m* and *ln m* are significant predictors of the cross-section of returns. However, the results for non-NASDAQ stocks do not support significance of idiosyncratic risk, consistent with the underlying economic model that predicts that premium for more widely-followed securities should be lower or insignificant; however, the small number of portfolios invites further analysis of those results.

#	Filter	<i>Beta</i>	<i>ln(Cap)</i>	<i>ln(B/M)</i>	<i>Ret(-7, -2)</i>	<i>Roll</i>	<i>m</i>	<i>ln m</i>	<i>R</i> ²
1	All	1.35 (2.86)	0.02 (0.48)	0.52 (3.11)					37.39
2	All	1.10 (2.47)	0.04 (1.13)	0.66 (4.88)	0.00 (0.89)				40.95
3	All	0.47 (1.02)					0.04 (2.16)		28.98
4	All	-0.05 (-0.10)						0.66 (2.74)	28.95
5	All	0.77 (1.71)	0.06 (1.76)	0.67 (4.65)			0.04 (2.34)		41.93
6	All	0.15 (0.34)	0.08 (1.93)	0.66 (4.39)				0.72 (3.31)	41.35

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	<i>m</i>	$\ln m$	R^2
7	All	0.55 (1.30)	0.07 (2.15)	0.71 (5.52)	0.00 (0.40)		0.04 (2.67)		44.81
8	All	0.00 (0.01)	0.09 (2.42)	0.74 (5.67)	0.00 (0.66)			0.69 (3.33)	44.36
9	UP>10	0.28 (0.67)	0.06 (1.65)	0.63 (5.02)	0.01 (1.14)		0.04 (2.38)		42.27
10	UP>10	-0.11 (-0.27)	0.08 (1.97)	0.70 (5.47)	0.01 (1.36)			0.58 (3.00)	41.96
11	non-NASDAQ	0.61 (1.38)	0.04 (0.94)	0.78 (6.13)	0.01 (1.66)		0.03 (2.09)		38.30
12	non-NASDAQ	0.18 (0.42)	0.05 (1.22)	0.79 (6.23)	0.01 (1.52)			0.51 (2.65)	38.20
13	All	0.61 (1.46)	0.07 (2.20)	0.66 (5.36)	0.00 (0.83)	0.02 (0.36)	0.04 (1.63)		47.25
14	All	0.07 (0.18)	0.07 (2.01)	0.67 (5.55)	0.00 (0.51)	0.00 (-0.01)		0.67 (2.69)	46.98
15	UP>10	0.37 (0.96)	0.06 (1.60)	0.59 (4.60)	0.01 (1.57)	-0.01 (-0.16)	0.04 (1.59)		44.59
16	UP>10	0.00 (-0.01)	0.07 (1.86)	0.64 (5.04)	0.01 (1.54)	-0.06 (-0.89)		0.74 (3.03)	44.28

#	Filter	<i>Beta</i>	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	<i>Roll</i>	<i>m</i>	$\ln m$	R^2
17	non-NASDAQ	0.40 (0.95)	0.04 (0.96)	0.69 (5.59)	0.01 (2.10)	0.10 (1.90)	0.00 (0.03)		40.69
18	non-NASDAQ	0.19 (0.47)	0.04 (0.94)	0.68 (5.64)	0.01 (1.66)	0.07 (1.34)		0.22 (0.87)	40.69

Source: author's calculations

The estimates from the Fama–MacBeth regressions using portfolios as assets are reported in Table 24. Specifications 3 through 16 confirm the conclusions from the preceding section, with idiosyncratic risk found to be a significant predictor of the cross-section of portfolio returns. Even in specifications 13 and 15, where the t -statistic for the idiosyncratic risk variable is below the 90% confidence level, the sign and magnitude of the coefficient are unchanged.

An interesting exception emerges in specifications 17 and 18, where idiosyncratic risk is found to be an insignificant predictor of the cross-section. One possible explanation could be the fairly high number of correlated predictors that results in difficulties in separating the contributions of the individual covariates. However, our deductive approach suggests that we should look at the underlying economic theory for clues, before drawing conclusions from these negative results. We believe that the selection of explanatory variables should be motivated by underlying economic theories and not by data mining. The baseline model for the significance of idiosyncratic risk is Merton (1987); it predicts that for a given beta and size, equilibrium alphas increase with the decrease of the share of investors that follow that security. In order to explore whether that is consistent with our data we form five size-quintile portfolios using the NYSE breakpoints, and then split each quintile into five quintile portfolios based on the mean-reverting level of idiosyncratic volatility, again using the NYSE breakpoints. The alphas relative to the Fama–French–Carhart model are reported in Table 25; in parentheses we report the t -statistics calculated using Newey and West (1987) standard errors.

Table 25: Portfolio Alphas Relative to Fama–French–Carhart Model

The table reports portfolio alphas relative to Fama-French-Carhart model for 25 double-sorted portfolios formed by market capitalisation ('Size') and mean-reverting level of volatility. Newey-West t-statistics are reported in parantheses. 'Average' is the average alpha for given volatility bucket. '(H-L)' is the spread between alpha for highest-volatility and lowest-volatility stocks.

The results suggest that alpha generally increases with the increase of idiosyncratic volatility and decrease of size. The average spread between the high-volatility portfolio and the low-volatility portfolio comes at 0.61 per cent per month, an economically and statistically significant difference. Also consistent with the economic model, the premium for idiosyncratic risk increases as size decreases. A significant exception is the first size quintile, which contains broadly followed stocks with market capitalisation of about half of the entire market. In that size quintile there is only a small difference between the first two volatility quintiles, and the relationship flattens afterwards.

Size	Mean-reverting level of volatility					
	Low	2	3	4	High	(H-L)
Large	0.31	0.46	0.49	0.49	0.49	0.18
	(1.51)	(1.94)	(1.88)	(1.82)	(1.38)	(0.77)
2	0.29	0.63	0.80	0.51	0.75	0.46
	(1.47)	(2.48)	(2.83)	(1.77)	(1.99)	(1.80)
3	0.39	0.63	0.67	0.79	0.71	0.32
	(1.90)	(2.47)	(2.31)	(2.61)	(1.79)	(1.16)
4	0.53	0.73	0.77	0.83	0.98	0.46
	(2.45)	(2.63)	(2.60)	(2.62)	(2.42)	(1.74)
Small	0.44	0.82	0.85	1.26	1.31	0.87
	(1.74)	(3.00)	(2.95)	(3.95)	(3.59)	(3.85)
Average	0.39	0.67	0.74	0.89	1.01	0.61
	(1.93)	(2.69)	(2.70)	(3.07)	(2.76)	(2.67)

Source: author's calculations

Consistent with other studies, Table 25 shows that alpha increases with the increase of idiosyncratic volatility and decrease of size. The average spread between the high-volatility portfolio and the low-volatility portfolio comes at 0.61 per cent per month, an economically and statistically significant difference. The two exceptions are the first and the third size quintiles. In the case of the latter, the result is driven mostly by the somewhat lower alpha for the high-volatility portfolio, while the middle four portfolios demonstrate a gradual increase of alpha with the increase of volatility, consistent with the hypothesis. Similarly and consistent with the economic model, the premium for idiosyncratic risk increases as size decreases. The significant exception is the first size quintile where there is only a small difference between the first two volatility quintiles, and the relationship flattens afterwards.

This seems particularly important given that those securities account for more than half of the whole market (in our sample). Again, it is important to recognise that this finding does not contradict the underlying economic model. The size of the alpha premium that accrues is not fixed but depends on how many investors actually track the securities in those portfolios. Therefore, even though the spread between the mean volatility of high-volatility and small-volatility shares for the first two size quintiles is similar (6 percentage points (p.p.) for the largest shares vs 7.4 p.p. for the second size quintile), the difference in alphas increases from 0.18 p.p. to 0.46 p.p. Such increases of alphas for a given volatility quintile are observed in the majority of cases. It seems reasonable to assume that the largest shares in terms of capitalisation are also more widely followed and consequently their premia for assuming idiosyncratic risk are smaller compared to the less-followed securities in the smaller size quintiles. Similarly, for a fixed series length smaller spreads in alphas would make rejection of the null hypothesis more unlikely. In that vein, the standard errors of the difference between the alphas of high-volatility and low-volatility stocks are very similar for the five size quintiles and the decision whether or not to reject the null hypothesis is driven mostly by the magnitude of the spread in alphas, which by the previous argument would be smaller for the better-known large stocks. Therefore, the mixed evidence reported by portfolio studies like Bali and Cakici (2008) as well as our results in this section could be seen as supporting the underlying model rather than contradicting it.

The table also shows that the alphas for the lowest-volatility stocks are materially below those of medium- and high-volatility stocks. Indeed, our table suggests that the steepest increase of alphas occurs between the lowest-volatility quintile portfolio and the second quintile portfolio. From second quintile until fifth quintile portfolio the alphas are either fairly flat (esp. for the largest cap quintile), or increase less steeply (the second and third size quintiles). Thus, the table does not exhibit the anomalies documented by Li et al. (2014)¹⁴⁶ and Walkshäusl (2013), who explore the finding of Ang et al. (2009, 2006) that there is a negative relationship between idiosyncratic volatility and expected returns. Walkshäusl explains such negative link between volatility and returns in terms of quality premium for low-volatility firms, where securities with more stable cash flows are concentrated. Li et al. (2014) point out that the strategy of zero investment portfolio that is long in low-volatility stocks and short in high-volatility stocks could be difficult to exploit due to the low liquidity

¹⁴⁶ Note that their study performs the sorts based on the Ang volatility measure, and should be interpreted in the context of return reversals for high-volatility stocks and the stationarity of volatility series, both of which are documented by Fu (2009).

and high cost of implementation. When portfolio sorts are implemented based on mean-reverting volatilities, we find no low-volatility anomaly (low-volatility stocks earning higher return than higher-volatility stocks), which suggests that the anomalies reported by other studies are transient phenomena that are due to random deviations of volatilities from their mean levels (e.g. due to arrival of new information). Therefore, if there is a puzzle, it would be why the difference of yields between the lowest-volatility quintile and the second-lowest volatility quintiles is so pronounced. At any rate, such low-volatility firms tend to have significantly lower betas and potentially – more stable earnings, although the latter does not necessarily translate into higher returns.¹⁴⁷

4.5.4. Interaction effects

Thus far we examined how the different measures of expected idiosyncratic volatility explained the cross-section of returns. We found that the estimators that yielded best one-period-ahead forecasts (ARMA and Ang’s IV_{t-1}) in fact were of limited use in explaining the cross-section. Then we saw that the mean-reverting level of volatility was a robust explanatory variable, which remained significant across various breakdowns of the sample. Those tests share one caveat with most of the existing studies – the tested models arguably does not accurately reflect the predictions of the tested economic model. Specifically, when reviewing the predictions of Merton’s model in section “2.2. Underlying economic theories” (p. 21) we noted that it predicted that the alpha earned by higher idiosyncratic risk would depend on two factors: the share of investors that know the security, and the variance of idiosyncratic shocks. In particular, for securities with given beta Merton (1987) showed (eq. 31.a on p. 496) that:

$$\frac{\partial \alpha_k}{\partial \sigma_k^2} = \delta(1 - q_k)x_k/q_k > 0, \quad (15)$$

where α_k is the alpha earned for investing in company k , σ_k^2 is the exposure to idiosyncratic risk, $q_k \in (0,1]$ is the share of investors that ‘know’ security k , δ is the parameter of the quadratic preference function, and x_k is the share of the value of company k in the overall market capitalisation. Thus for given beta and given company size x_k , securities that are in the investment set of all investors ($q_k = 1$) would earn no premium for idiosyncratic risk, while for companies that are not followed by all investors ($q_k < 1$), lower share of investors investing in that security would translate into higher alpha for idiosyncratic

¹⁴⁷ For arguments against the link between stable earnings and expected returns see Chapter 5 in Damodaran (2004)

risk. This non-linear relationship between α_k and q_k and σ_k^2 is not taken into account in the preceding tests, which effectively assume that $\partial\alpha_k/\partial\sigma_k^2 = \text{const}$, i.e. $q_k = \text{const}$ and $x_k = \text{const}$ for all k , which was a counter-factual assumption. Therefore, the tests in the preceding section might suffer from a functional form misspecification problem, which could result in biased estimates.

Similarly, Merton derived the comparative static with respect to size (x_k) and investor recognition (q_k):

$$\frac{\partial\alpha_k}{\partial x_k} = \frac{\delta(1-q_k)\sigma_k^2}{q_k} > 0,$$

$$\frac{\partial\alpha_k}{\partial q_k} = -\frac{\delta x_k \sigma_k^2}{q_k^2}.$$

However, the signs of the partial derivatives do not guarantee the sign of the total derivatives. In particular, since size correlates with volatility and investor recognition, its impact on alpha is unclear. Merton points out that $\frac{d\alpha_k}{dx_k} = \frac{\partial\alpha_k}{\partial x_k} + \frac{\partial\alpha_k}{\partial\sigma_k^2} \frac{d\sigma_k^2}{dx_k} + \frac{\partial\alpha_k}{\partial q_k} \frac{dq_k}{dx_k}$, the sign of which is uncertain and in principle could be negative (i.e. $d\alpha_k/dx_k < 0$) even though the partial derivative is positive ($\partial\alpha_k/\partial x_k > 0$). The same point would in principle apply to the total derivative with respect to volatility, i.e. $\frac{d\alpha_k}{d\sigma_k^2} = \frac{\partial\alpha_k}{\partial\sigma_k^2} + \frac{\partial\alpha_k}{\partial x_k} \frac{dx_k}{d\sigma_k^2} + \frac{\partial\alpha_k}{\partial q_k} \frac{dq_k}{d\sigma_k^2}$, and again if larger firms have lower volatilities so that $\frac{dx_k}{d\sigma_k^2} < 0$, then the alpha on size could in principle offset the partial derivative with respect to volatility.

In this section we shall extend our study and examine if idiosyncratic risk interacts with other explanatory variables in the cross-section. We see that the right-hand side is a linear function of company capitalisation (for a given market size) and a non-linear function of q_k . Unfortunately, we do not have information on q_k , the share of investors knowing each specific share. However, there are other proxy variables which may correlate with the share of investors following a given security. Such “instrument variables” could be the capitalisation of each company, the unadjusted price, and liquidity as measured by Roll’s bid-ask spread and the trade volume in each security. Concerning unadjusted price we note that previous research suggested that institutional investors prefer securities with higher price, e.g. over *USD* 10, in order to reduce transaction costs. Another characteristic that could be useful here is again the market capitalisation, where one could speculate that companies with large capitalisation are better known to investors and more widely tracked than small-cap stocks. Furthermore, liquidity, as measured by the Roll’s bid-ask spread and by the traded value in

the last 24 to 36 month (as available), are also likely correlated with idiosyncratic volatility.

We examine possible differences in idiosyncratic risk slopes as follows: at each month we split our sample into four quartile portfolios based on some variable that could be correlated with idiosyncratic risk; we then estimated the cross-sectional regression using only the observations in that quartile portfolio.¹⁴⁸ Thus, for each specification we obtain four estimates – one for each quartile of the control variable. We employ three such control variables: capitalisation (*Cap*)¹⁴⁹, liquidity, as measured by the Roll's estimator of the bid-ask spread (*Roll*)¹⁵⁰, the unadjusted price¹⁵¹, and traded turnover.¹⁵² This allows us to examine if significance of idiosyncratic volatility is limited only to some specific subsets of the cross-section of returns, and whether it changes with the instrument variables. The split by capitalisation is aimed to capture differences in size, and in particular whether idiosyncratic risk is a significant predictor of the cross-section among larger companies. The splits by Roll's bid-ask spread and by traded value are intended to test whether idiosyncratic risk matters only for the most illiquid stocks. Finally, the split by unadjusted prices is aimed to test if idiosyncratic volatility mattered mostly for low-price stocks, which could be associated with financial distress.

We should also note that the tests in this section share one drawback: since the split into quartile segments is based on variables presumably correlated with idiosyncratic volatility, within each quartile the variation of idiosyncratic volatility is reduced, which may render more difficult the rejection of the null hypothesis and result in seemingly insignificant coefficients of idiosyncratic volatility and other covariates, correlated with the instrument. In such cases, the introduction of three dummy variables for three of the four quartiles could allow us to capture changes in mean alphas within each segment; however, it would be difficult to attribute the coefficients of these dummies to any specific factor among the ones correlated with idiosyncratic risk. Therefore, we focus on the baseline methodology outlined in preceding paragraphs of this section, but the reader should interpret indications of insignificance of some coefficients with a grain of salt.

¹⁴⁸ Another approach could be to incorporate directly interaction terms in the tested specification. That approach could work more easily for variable that is already present in the specification, like size and Roll's bid-ask spread, but is more difficult to justify when we used other control variables like unadjusted price or traded volume.

¹⁴⁹ see Table 26 on p. 225

¹⁵⁰ see Table 27 on p. 228

¹⁵¹ see Table 28 on p. 231

¹⁵² see Table 29 on p. 234

Table 26 summarises the tests of idiosyncratic risk by quartiles of market capitalisation. In all four quartiles we observe similar statistically-significant coefficients for the mean-reverting volatility. In fact, contrary to what we anticipated, the coefficients turned out positive and significant even in the high-capitalisation quartile, and in fact the point estimate is slightly higher than the corresponding estimate for the small-cap quartile. In part this may be due to more noise in the small-cap quartile. An indication in that direction is the generally lower values of adjusted- R^2 for the small-cap securities compared to large-cap ones; this pattern we observe also in some of the other analyses from this section.

The patterns documented in Table 27 are similar to those reported above: the coefficients for idiosyncratic volatilities are numerically similar and significant across Roll's bid-ask quartiles, and again the point estimate for the low-spread (high-liquidity) quartile is in fact somewhat higher than the corresponding value for the high-spread (low-liquidity) quartile, which also has a somewhat lower adjusted- R^2 coefficient.

Table 28 summarises the interaction between idiosyncratic volatilities and unadjusted prices. Again, contrary to our expectations we find that idiosyncratic volatilities are a significant predictor even in the high-price quartile, where investors could be expected to hold more diversified portfolios and require smaller risk premia for idiosyncratic risk. Again the coefficient estimates are not too dissimilar across quartiles, but high-price quartiles show higher point estimate and higher adjusted- R^2 coefficient compared to the low-price quartile.

A notable exception from the above results is the interaction between the traded volume and idiosyncratic risk premium. Among all controls tested in this section, this one is perhaps most directly related to the underlying concept of breadth of investors 'knowing' a security. Indeed, securities that are traded consistently in high volumes are likely to be well-known to investors and should earn no or a negligible risk premium. The results in Table 29 are broadly consistent with the model predictions: in the lowest-traded volume quartile we found significant correlation between idiosyncratic risk and the cross-section of returns. Surprisingly, however, the coefficients for the remaining three quartiles are insignificant; clearly, we expected that it should be insignificant in at least the highest-traded-volume quartile and possibly in the second-highest quartile, but we find that it is insignificant for the subsamples that covered 75% of our sample. The point estimates, however, remained remarkably stable within the three quartiles, which could be consistent with the limitation of the methodology that we outlined earlier in this section, viz. that the null hypothesis may be difficult to reject due to the limited variability of idiosyncratic volatility in each of the quartiles. We have therefore explored this further in specifications 25-30 in Table 29, where

we pool together the three quarters of securities with higher traded volume. In those specifications we again find significant correlation between idiosyncratic risk and returns, which supports the view that the insignificant coefficients in some of the specifications 7 through 24 could be driven by the reduced variance of the explanatory variable, rather than by its insignificance per se.

Table 26: Fama–MacBeth regressions with mean-reverting level of volatility - interaction with size, 7/1980–3/2013

The table reports results from Fama-Macbeth cross-sectional regressions for subsets formed by size (market capitalisation). In each month all stocks are sorted in four quartile groups based on the capitalisation and the cross-sectional regressions are estimated separately for each quartile group.

Control variables are beta with market (*Beta*), natural logarithms of market capitalisation (*ln(Cap)*) and book-to-market value (*ln(B/M)*), momentum of returns measured as cumulative return from (t-7) until (t-1) (*Ret(-7, -2)*), return in the previous month (*Ret_{t-1}*), liquidity measured in terms of Roll's estimator (*Roll*). Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. *R*² reports the averaged R-squared statistics from the cross-sectional regressions.

Column 'Filter' indicates the subsample used for estimating the cross-sectional regressions: 'high-cap' marks Fama-Macbeth regressions estimated using only 25% of stocks with highest capitalisation in each month. '2nd quartile', '3rd quartile' and 'small-cap' mark the remaining three quartile groups.

The results suggest that *m* and *ln m* are significant predictors of the cross-section of returns for all quartile groups. However, the slope does not appear to increase with capitalisation, contrary to expectations.

#	Filter	<i>Beta</i>	<i>ln(Cap)</i>	<i>ln(B/M)</i>	<i>Ret(-7, -2)</i>	<i>Roll</i>	<i>Ret_{t-1}</i>	<i>m</i>	<i>ln m</i>	<i>R</i> ²
1	high-cap	0.23 (0.86)						0.01 (0.55)		5.60
2	high-cap	0.20 (0.75)	0.02 (0.41)	0.33 (3.27)				0.03 (1.27)		7.17
3	high-cap	0.14 (0.60)	0.02 (0.36)	0.33 (3.96)	0.00 (1.23)	-0.05 (-0.95)	-0.02 (-3.58)	0.04 (2.28)		10.99
4	high-cap	0.19 (0.72)							0.17 (0.74)	5.54
5	high-cap	0.14 (0.54)	0.04 (0.77)	0.34 (3.40)					0.38 (1.62)	7.14
6	high-cap	0.09 (0.42)	0.04 (0.86)	0.34 (4.20)	0.00 (1.27)	-0.07 (-1.18)	-0.02 (-3.59)		0.53 (3.08)	10.98

#	Filter	<i>Beta</i>	$\ln(Cap)$	$\ln(B/M)$	$Ret(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
7	2nd quartile	0.26 (0.94)						0.01 (0.53)		3.65
8	2nd quartile	0.24 (0.89)	0.01 (0.16)	0.72 (6.34)				0.05 (2.62)		4.57
9	2nd quartile	0.23 (0.97)	-0.03 (-0.36)	0.72 (7.74)	0.01 (2.44)	-0.07 (-1.46)	-0.03 (-5.21)	0.05 (3.61)		6.89
10	2nd quartile	0.17 (0.63)							0.25 (0.98)	3.65
11	2nd quartile	0.12 (0.44)	0.05 (0.61)	0.75 (6.43)					0.74 (3.32)	4.60
12	2nd quartile	0.13 (0.54)	0.00 (0.03)	0.75 (7.95)	0.01 (2.42)	-0.09 (-2.02)	-0.03 (-5.24)		0.87 (5.07)	6.91
13	3rd quartile	0.72 (2.31)						0.00 (-0.18)		2.39
14	3rd quartile	0.69 (2.27)	0.15 (1.20)	1.09 (10.45)				0.04 (2.88)		3.26
15	3rd quartile	0.61 (2.37)	0.16 (1.35)	1.16 (12.99)	0.01 (5.99)	0.01 (0.25)	-0.03 (-5.79)	0.03 (2.15)		5.24
16	3rd quartile	0.65 (2.16)							0.05 (0.21)	2.41

#	Filter	<i>Beta</i>	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	<i>Roll</i>	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
17	3rd quartile	0.59 (2.00)	0.18 (1.38)	1.10 (10.48)					0.74 (3.17)	3.30
18	3rd quartile	0.56 (2.19)	0.18 (1.44)	1.17 (13.04)	0.01 (6.03)	0.00 (0.00)	-0.03 (-5.77)		0.49 (2.66)	5.25
19	small-cap	0.72 (2.06)						0.03 (2.00)		1.23
20	small-cap	0.90 (2.67)	-0.05 (-0.49)	1.26 (10.34)				0.05 (3.95)		2.20
21	small-cap	0.62 (2.02)	0.15 (1.29)	1.23 (10.79)	0.01 (3.24)	0.06 (2.10)	-0.06 (-9.45)	0.03 (2.34)		4.08
22	small-cap	0.65 (1.89)							0.61 (2.29)	1.24
23	small-cap	0.81 (2.45)	-0.04 (-0.36)	1.26 (10.49)					1.07 (4.42)	2.21
24	small-cap	0.56 (1.86)	0.16 (1.34)	1.24 (10.94)	0.01 (3.31)	0.06 (1.95)	-0.06 (-9.46)		0.66 (3.02)	4.08

Source: author's calculations

Table 27: Fama–Macbeth regressions with mean-reverting level of volatility – interactions with Roll's bid-ask spread, 7/1980-03/2013

The table reports results from Fama-Macbeth cross-sectional regressions for subsets formed by liquidity (bid-ask spread). In each month all stocks are sorted in four quartile groups based on the bid-ask spread and the cross-sectional regressions are estimated separately for each quartile group.

Control variables are beta with market (*Beta*), natural logarithms of market capitalisation ($\ln(\text{Cap})$) and book-to-market value ($\ln(B/M)$), momentum of returns measured as cumulative return from (t-7) until (t-1) ($\text{Ret}(-7, -2)$), return in the previous month (Ret_{t-1}), liquidity measured in terms of Roll's estimator (*Roll*). Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. R^2 reports the averaged R-squared statistics from the cross-sectional regressions.

Column 'Filter' indicates the subsample used for estimating the cross-sectional regressions: 'high-liquidity' marks Fama-Macbeth regressions estimated using only 25% of stocks with highest liquidity (lowest bid-ask spread) in each month. '2nd quartile', '3rd quartile' and 'low-liquidity' mark the remaining three quartile groups.

The results suggest that *m* and $\ln m$ are significant predictors of the cross-section of returns for all quartile groups, confirming that the result is not due to a small subset of illiquid stocks and is therefore tradable even for the most liquid stocks.

#	Filter	Beta	$\ln(\text{Cap})$	$\ln(B/M)$	$\text{Ret}(-7, -2)$	Roll	Ret_{t-1}	<i>m</i>	$\ln m$	R^2
1	high-liquidity	0.36 (1.39)						0.06 (3.36)		3.31
2	high-liquidity	0.30 (1.15)	0.04 (1.42)	0.20 (2.85)				0.07 (4.10)		5.07
3	high-liquidity	0.25 (1.04)	0.03 (1.14)	0.18 (2.84)	0.01 (1.85)	0.04 (0.70)	-0.05 (-7.75)	0.06 (3.55)		7.38
4	high-liquidity	0.30 (1.17)							0.53 (3.88)	3.41
5	high-liquidity	0.22 (0.87)	0.06 (1.98)	0.21 (3.10)					0.73 (4.79)	5.19
6	high-liquidity	0.20 (0.84)	0.05 (1.71)	0.19 (3.11)	0.01 (1.85)	0.01 (0.19)	-0.05 (-7.79)		0.65 (4.33)	7.43

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7,-2)	Roll	Ret _{t-1}	m	ln m	R ²
7	2nd quartile	0.15 (0.67)						0.05 (3.62)		1.51
8	2nd quartile	0.16 (0.73)	0.04 (1.26)	0.56 (5.61)				0.06 (3.96)		3.14
9	2nd quartile	0.07 (0.32)	0.04 (1.23)	0.56 (6.63)	0.01 (2.42)	0.03 (0.38)	-0.04 (-8.08)	0.05 (3.44)		6.24
10	2nd quartile	0.11 (0.47)							0.62 (3.99)	1.57
11	2nd quartile	0.10 (0.46)	0.07 (1.90)	0.58 (5.86)					0.79 (4.66)	3.17
12	2nd quartile	0.02 (0.08)	0.06 (1.78)	0.58 (6.84)	0.01 (2.40)	0.00 (0.04)	-0.04 (-8.12)		0.69 (4.19)	4.84
13	3rd quartile	0.28 (1.03)						0.03 (1.97)		1.29
14	3rd quartile	0.47 (1.90)	0.07 (1.20)	1.00 (8.96)				0.04 (3.04)		3.06
15	3rd quartile	0.40 (1.68)	0.05 (0.85)	1.07 (11.13)	0.01 (5.41)	0.02 (0.37)	-0.03 (-5.69)	0.03 (2.51)		4.52
16	3rd quartile	0.25 (0.94)							0.44 (2.00)	1.34
17	3rd quartile	0.41	0.10	1.03					0.81	3.12

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
		(1.70)	(1.62)	(9.37)					(3.37)	
18	3rd quartile	0.35	0.07	1.10	0.01	0.01	-0.03		0.66	4.56
		(1.51)	(1.23)	(11.52)	(5.38)	(0.17)	(-5.74)		(3.08)	
19	small-liquidity	0.14						0.03		1.07
		(0.37)						(2.23)		
20	small-liquidity	0.86	-0.07	1.35				0.04		2.89
		(2.59)	(-0.88)	(10.59)				(3.00)		
21	small-liquidity	0.80	0.00	1.40	0.01	0.04	-0.04	0.03		4.43
		(2.51)	(0.02)	(11.83)	(3.73)	(1.38)	(-8.61)	(1.94)		
22	small-liquidity	0.10							0.58	1.63
		(0.28)							(2.03)	
23	small-liquidity	0.82	-0.06	1.37					0.84	2.90
		(2.48)	(-0.71)	(10.91)					(2.93)	
24	small-liquidity	0.76	0.01	1.40	0.01	0.04	-0.04		0.57	4.45
		(2.40)	(0.16)	(12.11)	(3.75)	(1.37)	(-8.64)		(2.01)	

Source: author's calculations

Table 28: Fama–Macbeth regressions with mean-reverting level of volatility – interaction with unadjusted prices, 07/1980-03/2013

The table reports results from Fama-Macbeth cross-sectional regressions for subsets formed by unadjusted prices. In each month all stocks are sorted in four quartile groups based on the unadjusted price and the cross-sectional regressions are estimated separately for each quartile group.

Control variables are beta with market (*Beta*), natural logarithms of market capitalisation (*ln(Cap)*) and book-to-market value (*ln(B/M)*), momentum of returns measured as cumulative return from (t-7) until (t-1) (*Ret(-7, -2)*), return in the previous month (*Ret_{t-1}*), liquidity measured in terms of Roll’s estimator (*Roll*). Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. *R*² reports the averaged R-squared statistics from the cross-sectional regressions.

Column ‘Filter’ indicates the subsample used for estimating the cross-sectional regressions: ‘low-price’ marks Fama-Macbeth regressions estimated using only 25% of stocks with lowest unadjusted price in each month. ‘2nd quartile’, ‘3rd quartile’ and ‘high-price’ mark the remaining three quartile groups.

The results suggest that *m* and *ln m* are significant predictors of the cross-section of returns for all quartile groups. The results suggest that significance of idiosyncratic volatility is not due to subset of small-price stocks that are avoided by institutional investors due to transaction costs. On the other hand, there is no clear trend in the regression slopes. Low-priced stocks would be preferred by individual investors who are less diversified, so *ceteris paribus* should also earn higher premium for assuming idiosyncratic risk.

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
1	low-price	0.35 (1.00)						0.02 (1.22)		1.38
2	low-price	0.66 (2.00)	0.01 (0.10)	1.31 (11.55)				0.05 (3.71)		2.58
3	low-price	0.51 (1.64)	0.10 (1.22)	1.31 (12.50)	0.01 (2.76)	0.07 (2.53)	-0.05 (-9.35)	0.03 (2.06)		4.28
4	low-price	0.29 (0.84)							0.33 (1.05)	1.43
5	low-price	0.58 (1.79)	0.03 (0.42)	1.33 (11.97)					1.01 (3.80)	2.61
6	low-price	0.45 (1.47)	0.12 (1.47)	1.33 (12.87)	0.01 (2.81)	0.07 (2.50)	-0.05 (-9.41)		0.60 (2.44)	4.29

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
7	2nd quartile	0.29 (0.95)						0.00 (0.24)		2.50
8	2nd quartile	0.43 (1.53)	0.09 (1.82)	0.97 (9.51)				0.04 (2.57)		3.68
9	2nd quartile	0.46 (1.72)	0.10 (2.04)	1.01 (11.46)	0.01 (5.91)	-0.07 (-1.94)	-0.03 (-5.94)	0.03 (2.74)		5.43
10	2nd quartile	0.23 (0.77)							0.13 (0.54)	2.53
11	2nd quartile	0.33 (1.21)	0.12 (2.34)	1.00 (9.81)					0.72 (3.20)	3.74
12	2nd quartile	0.37 (1.42)	0.13 (2.58)	1.04 (11.71)	0.01 (5.93)	-0.08 (-2.48)	-0.03 (-5.93)		0.71 (3.91)	5.46
13	3rd quartile	0.24 (0.98)						0.01 (0.57)		3.81
14	3rd quartile	0.34 (1.37)	-0.01 (-0.33)	0.57 (5.77)				0.03 (1.83)		5.02
15	3rd quartile	0.33 (1.52)	-0.01 (-0.34)	0.57 (6.94)	0.01 (3.26)	-0.07 (-1.78)	-0.03 (-6.06)	0.03 (2.25)		7.16
16	3rd quartile	0.15 (0.65)							0.25 (1.11)	3.81

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
17	3rd quartile	0.23 (0.98)	0.01 (0.24)	0.59 (5.97)					0.54 (2.63)	5.03
18	3rd quartile	0.25 (1.19)	0.01 (0.31)	0.60 (7.16)	0.01 (3.26)	-0.09 (-2.27)	-0.03 (-6.04)		0.59 (3.56)	7.15
19	high-price	0.20 (0.82)						0.04 (1.47)		4.33
20	high-price	0.17 (0.71)	0.04 (1.10)	0.36 (4.25)				0.06 (2.49)		5.85
21	high-price	0.17 (0.82)	0.03 (1.01)	0.38 (4.74)	0.01 (1.96)	-0.05 (-0.96)	-0.02 (-3.90)	0.06 (3.64)		8.84
22	high-price	0.16 (0.68)							0.34 (1.38)	4.29
23	high-price	0.11 (0.45)	0.06 (1.55)	0.38 (4.40)					0.65 (2.52)	5.85
24	high-price	0.12 (0.60)	0.06 (1.63)	0.39 (4.94)	0.01 (2.03)	-0.06 (-1.18)	-0.02 (-3.82)		0.69 (4.22)	8.84

Source: author's calculations

Table 29: Fama–Macbeth regressions with mean-reverting level of volatilities – interaction with traded volume in the last 36 months, 07/1980-03/2013

The table reports results from Fama-Macbeth cross-sectional regressions for subsets formed by traded volume. In each month all stocks are sorted in four quartile groups based on the traded volume and the cross-sectional regressions are estimated separately for each quartile group.

Control variables are beta with market (*Beta*), natural logarithms of market capitalisation (*ln(Cap)*) and book-to-market value (*ln(B/M)*), momentum of returns measured as cumulative return from (t-7) until (t-1) (*Ret(-7, -2)*), return in the previous month (*Ret_{t-1}*), liquidity measured in terms of Roll’s estimator (*Roll*). Tested variables are the mean-reverting level of volatility (*m*) and its natural logarithm. *R*² reports the averaged R-squared statistics from the cross-sectional regressions.

Column ‘Filter’ indicates the subsample used for estimating the cross-sectional regressions: ‘low-volume’ marks Fama-Macbeth regressions estimated using only 25% of stocks with lowest traded volume in each month. ‘2nd quartile’, ‘3rd quartile’ and ‘high-volume’ mark the remaining three quartile groups. ‘75% volume’ pools together the stocks from the top 75% in terms of traded volume.

The results suggest that *m* and *ln m* are significant predictors of the cross-section of returns only for stocks with low traded volume. However, the remaining point estimates remain notably stable although statistically insignificant. In order to examine if the insignificance of the coefficient is due to some specific months, e.g. ones with high volatility, we have pooled together the top 75% into one portfolio, which shows a statistically significant slope, confirming the significance of the mean-reverting volatility.

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
1	low-volume	0.26 (0.83)						0.05 (3.39)		1.72
2	low-volume	0.33 (1.12)	-0.18 (-2.89)	1.05 (9.87)				0.05 (3.72)		2.90
3	low-volume	0.22 (0.82)	-0.10 (-1.69)	1.06 (10.30)	0.01 (5.49)	0.05 (2.09)	-0.05 (-9.13)	0.03 (2.26)		4.52
4	low-volume	0.17 (0.58)							0.83 (3.69)	1.72
5	low-volume	0.28 (0.97)	-0.17 (-2.79)	1.04 (9.93)					0.82 (3.84)	2.91

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
6	low-volume	0.19 (0.73)	-0.10 (-1.61)	1.05 (10.40)	0.01 (5.53)	0.05 (2.20)	-0.05 (-9.16)		0.41 (2.44)	4.52
7	2nd quartile	0.64 (2.08)						0.04 (2.17)		3.27
8	2nd quartile	0.38 (1.36)	-0.43 (-4.71)	0.97 (9.93)				0.02 (1.57)		4.14
9	2nd quartile	0.25 (1.00)	-0.36 (-4.77)	1.02 (11.35)	0.01 (5.81)	0.00 (0.05)	-0.04 (-7.99)	0.02 (1.33)		6.03
10	2nd quartile	0.51 (1.75)							0.63 (2.53)	2.80
11	2nd quartile	0.31 (1.19)	-0.41 (-4.35)	0.97 (9.94)					0.42 (2.01)	4.13
12	2nd quartile	0.21 (0.86)	-0.35 (-4.47)	1.03 (11.34)	0.01 (5.88)	0.00 (-0.02)	-0.04 (-8.04)		0.28 (1.76)	6.02
13	3rd quartile	0.30 (1.07)						0.06 (2.70)		3.77
14	3rd quartile	0.07 (0.29)	-0.49 (-4.81)	0.75 (6.70)				0.01 (0.77)		5.40
15	3rd quartile	-0.03 (-0.15)	-0.49 (-6.68)	0.78 (8.01)	0.01 (3.81)	-0.07 (-1.45)	-0.03 (-5.49)	0.02 (1.18)		7.77

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
16	3rd quartile	0.17 (0.63)							0.73 (2.91)	3.75
17	3rd quartile	0.00 (0.01)	-0.46 (-4.46)	0.78 (6.78)					0.27 (1.19)	5.37
18	3rd quartile	-0.11 (-0.46)	-0.46 (-6.27)	0.80 (8.13)	0.01 (3.80)	-0.08 (-1.73)	-0.03 (-5.48)		0.38 (1.76)	7.78
19	high-volume	0.18 (0.66)						0.08 (3.22)		5.48
20	high-volume	0.04 (0.16)	-0.25 (-4.36)	0.42 (3.62)				0.03 (1.37)		7.27
21	high-volume	-0.02 (-0.08)	-0.24 (-4.66)	0.44 (4.58)	0.01 (1.85)	0.03 (0.48)	-0.01 (-2.52)	0.02 (1.08)		10.68
22	high-volume	0.12 (0.43)							0.81 (3.37)	5.44
23	high-volume	0.03 (0.13)	-0.25 (-4.22)	0.42 (3.61)					0.33 (1.32)	7.19
24	high-volume	-0.03 (-0.12)	-0.24 (-4.50)	0.44 (4.60)	0.01 (1.84)	0.03 (0.46)	-0.01 (-2.47)		0.21 (1.29)	10.61
25	75% volume	0.41 (1.50)						0.04 (2.51)		3.51

#	Filter	Beta	ln(Cap)	ln(B/M)	Ret(-7, -2)	Roll	Ret _{t-1}	m	ln m	R ²
26	75% volume	0.34 (1.30)	-0.04 -(0.96)	0.77 (7.50)				0.04 (2.70)		4.90
27	75% volume	0.19 (0.87)	-0.03 -(0.70)	0.78 (9.58)	0.01 (3.38)	0.01 (0.31)	-0.03 -(7.06)	0.03 (2.26)		7.04
28	75% volume	0.30 (1.17)							0.59 (2.77)	3.54
29	75% volume	0.24 (0.97)	-0.01 -(0.25)	0.79 (7.66)					0.66 (3.02)	4.94
30	75% volume	0.12 (0.56)	0.00 -(0.05)	0.80 (9.75)	0.01 (3.38)	0.00 (0.06)	-0.03 -(7.04)		0.49 (3.11)	7.07

Source: author's calculations

4.6. Summary

In this chapter we presented the empirical finding on the correlation between idiosyncratic risk and stock returns. We found that past volatilities and forecasts derived using ARMA(1,1) based on past volatilities substantially outperformed forecasts from monthly GARCH(1,1) or OLS residuals as predictors of next-period volatility. Nonetheless, we found that those superior forecasts did not yield support to the hypothesis that idiosyncratic risk is priced. In particular, we found that the empirical evidence in favour of negative correlation noted by Ang et al. (2006) and Ang et al. (2009) was fragile and could not be confirmed as a robust predictor of the cross-section after controlling for the skewness of the lagged volatility distribution. On the other hand, the forecasts from OLS and GARCH(1,1) did support the existence of such positive correlation, consistent with the findings of Fu (2009) and Spiegel and Wang (2005). In view of the negative results for one-period forecasts we explored the predictive performance of another key characteristic of expect volatility path – the mean-reverting level of volatility. We found that the mean-reverting level was a robust predictor of the cross-section after controlling for many characteristics and factor loadings, including beta, size, book-market ratio, return momentum, return reversals, liquidity, omitted factors, daily frequency, and various subsamples. In the following chapter we shall offer further interpretation of our results in order to understand how they fit with previous empirical and theoretical research.

5. Discussion

5.1. Introduction

The results in the preceding chapter convincingly demonstrated that next-period volatility does not explain the cross-section of stock returns, but the mean-reverting level of volatility is established as a robust predictor of the cross-section. In this chapter we shall step back and assess our findings against the backdrop of other studies and theoretical developments in the recent years. In Section 5.2. we shall discuss the findings concerning the superior performance of volatility forecasts based on daily data, and compare our conclusions with those reported by related studies. In Section 5.3. we shall examine the grounds and theories that motivate the use of the mean-reverting volatility as a predictor of the cross-section and basis for portfolio construction, instead of the short-term volatility forecasts. Section 5.4. offers an interpretation of the grounds for the conflicting conclusions reached by existing literature and shall demonstrate, that those studies are actually consistent with one another, as well as with the theoretical models. Section 5.5. is dedicated to the interaction of idiosyncratic volatility with other predictors of the cross-section, especially liquidity and size. There we explain why the idiosyncratic premium might not be separable from liquidity and size premia. Section 5.6. is dedicated to one domain where further research is needed, namely the tradability of idiosyncratic volatility.

5.2. Forecasts quality and goodness of fit

Existing studies employed different measures volatilities, which substantially obscured the comparison and interpretation of results. For example, Fu (2009) criticised the approach of Ang et al. (2006) as backward-looking rather than forward-looking, while Bali and Cakici (2008: 52) reported that monthly idiosyncratic volatility forecasts significantly outperformed daily volatility forecasts.

We approached the problem by starting off from the basic question which measure of volatility fared better in forecasting the true volatilities. We pursued that by using Mincer-Zarnowitz regressions, the results of which we reported in Table 13 on page 136. As predictor variables we used a couple of existing measures proposed in the literature: the OLS monthly rolling historical volatility, the last-month historical volatility used by Ang et al.

(2006), the expected volatility from the GARCH model, and the expected volatility from the ARMA model using monthly volatilities estimated from daily data. In this way we covered a broader range of possible measures without, however, ruling out the possibility of some methodological refinements, e.g. using Exponential GARCH as in Fu (2009) and Spiegel and Wang (2005) instead of simple GARCH. Nevertheless, in our empirical analysis we found a fairly wide gap in predictive performance between ARMA and Ang's forecasts, on the one hand, and monthly GARCH and OLS volatilities on the other hand, so we doubt that alternative specifications could close that gap, although we are confident that some specifications could improve predictive accuracy. Nevertheless, our result seem to be quite away from those of Spiegel and Wang, who reported¹⁵³ that the use of EGARCH(p,q) forecasts with lags p and q ranging between 1 and 3 reduced the forecasting error to just half of that of OLS residuals.

The separation of the predictive performance of idiosyncratic volatilities can be seen in terms of approaches to forecasting of volatilities. The approach of Ang et al. (2006) and its forward-looking extension using ARMA(1,1) were based on the average daily volatilities from the preceding month as a proxy of monthly volatility (after corresponding scaling). This approach was in the spirit of Merton (1980) in the sense that lower-frequency volatilities (in our case – monthly frequency) were estimated from higher-frequency data (daily data), which increased the accuracy (i.e., reduced the noise) of estimating monthly volatilities. The superior measures of past history of monthly volatilities then resulted in superior forecasts of next-month volatilities even though the approach of Ang et al. (2006) was essentially the simplest possible: it assumed that last month's volatility was a good proxy of present month's volatility. On the other hand, approaches based on monthly data used squared monthly return as a proxy of realised volatility. Thus, each month's volatility was estimated from a single realisation from the unobservable monthly return distribution, which in the present context was assumed to have time-varying scale. This was too big a hurdle to be overcome even with an otherwise powerful forecasting model like GARCH. The OLS residuals essentially reduced that problem by averaging those squared residuals over the rolling window, implicitly assuming a constant scale of the idiosyncratic return distributions. That would be a robust approach in case of static, unchanging monthly volatilities, but it filtered out most of the time variation of volatilities. The GARCH model addressed that problem by updating the previous forecast with the estimate of the realised volatility. However, the noisy update (squared

¹⁵³ Table 2 in Spiegel and Wang (2005)

residuals) was a significant handicap and capped the forecasting performance of the model. Nonetheless, the jury is still out, as Table 13 also demonstrated that the ranking of predictive models could also depend on the choice of estimator of the true volatility. Thus, the superiority of models using daily data to forecast monthly frequencies was significantly more pronounced when the realised volatility was measured using an in-sample EGARCH model on daily data, than the case where realised volatility equalled the squared idiosyncratic return. Therefore, different methods to estimate the true volatility could also impact the ranking, although we did not find evidence of that in our tests.

In our study we found that ARMA(1,1) yielded the most accurate predictor of future volatility among the compared alternatives. Using one-step forecasts from that model we found that idiosyncratic volatility appeared to be uncorrelated with returns. These results were reported in Table 15 on page 154. There we found that ARMA forecasts were insignificant across specifications both in levels and in logs. The historical estimator of Ang et al. (2006), which was the second best predictor of true volatilities in our Mincer and Zarnowitz (1969) tests, also yielded insignificant results once its skewness was reduced by using the natural logarithm of volatility. These results seemingly contradicted the theory of Merton that idiosyncratic risk was priced in case of imperfect diversification. On the other hand, the less accurate predictors in our study – GARCH and OLS – yielded support for the theory, as evident in Table 14 on page 148. The contrast between the results in Table 14 and Table 15 suggested that efforts to improve forecasting performance could be in fact leading us astray. If the better forecasts yielded inconclusive results, whereas the worse forecasts yielded stronger support for the theory of Levy and Merton, then we should conclude that those worse forecasts served as proxies for some other variable.

Our results cast new light on previous studies. Thus, we find that contrary to the findings of Bali and Cakici (2008), monthly idiosyncratic volatility was in fact an inferior predictor of future volatility. The difference in conclusions comes from their use of monthly data to estimate realised volatilities. When monthly volatilities were estimated from higher-frequency (daily) returns, the measures based on daily data (ARMA and Ang's measure) vastly outperformed those based on monthly data. The choice of estimator of the latent realised volatility was therefore found to affect significantly the conclusions of such comparisons. While the filtering of monthly volatilities from monthly returns may seem as more natural a choice, we point out that the short series and the theoretical works of Merton (1980) and Andersen and Bollerslev (1998) lend strong support for the use of higher-frequency data to estimate monthly volatilities. Using that approach we found that the

daily-based volatility forecasts outperformed monthly forecasts, and in particular that the use of GARCH with monthly data significantly underperformed as predictor of future volatility when compared to the historical volatilities of Ang et al. (2006). These findings suggest that the approach of Ang et al. (2006) and Ang et al. (2009) could not be discarded as a valid forecast of future volatilities.

5.3. Mean-reverting level

The analysis of the cause of the contrasting findings reported in Table 14, which supported the theory of Levy and Merton, and Table 15, which rejected it, should consider how volatilities change in time. In Table 12 on page 148 we reported that for the majority of companies the null hypothesis of unit root could be rejected at conventional confidence levels. This remained true when the true (*ex post*) volatility was measured by fitting the in-sample GARCH model with monthly data, as well as when we used a daily EGARCH model. The stationarity of the volatility series suggested that the level to which volatility reverted could be the one that actually mattered for investors. Indeed, the original model of Merton was a one-period model with known volatility. The model then derived the equilibrium prices given the known expected return and risk. In reality, investors faced a more complex decision problem: to estimate their allocations in the presence of changing volatilities. Thus, from the perspective of each investor, the return on asset i was determined by the sum of the expected dividend yield and the expected capital gains, i.e.

$$\mathbb{E}_t R_{i,t+1} = \frac{\mathbb{E}_t P_{i,t+1} - P_{i,t}}{P_{i,t}} + \frac{\mathbb{E}_t D_{i,t+1}}{P_{i,t}},$$

where $\mathbb{E}_t R_{i,t+1}$ was the expected return for the next period, $\mathbb{E}_t P_{i,t+1}$ was the expected price at the end of the allocation period, and $\mathbb{E}_t D_{i,t+1}$ was the expected dividend payment through the holding period. If equilibrium prices depended on the expected volatilities, as Merton proposed, then the equilibrium problem would need to change to incorporate the next-period volatility, as it would affect the end-of-period prices. Therefore, investors would be solving the problem backwards, starting from the unconditional mean (the mean-reverting level) and solving the problem backwards until the current level of volatility. Seen this way, the problem suggested a role of the unconditional mean as it determined the expected future trajectory of volatilities. On the other hand, such a perspective also supported the significance of the one-period volatility forecast, as this determined the distribution of the end-of-period prices,

and consequently, the end-of-period wealth. These two extremes – the start of the volatility trajectory, and its end – as well as the speed with which the volatility was expected to transit starting from the current level could all matter to investors. The volatility trajectories also mapped to the measures used in the empirical tests of the Merton model; thus, last-period volatility (i.e., the measure of Ang) could be seen as the starting point of the volatility at the time of allocation. The volatility during the investment period (the ARMA/GARCH one-period forecast) could be seen as an intermediary point on the trajectory linking the last volatility (Ang et al's) and the mean-reverting level of volatility.

The proposition that the mean-reverting level of volatility explains the cross-section could be linked to a number of previous studies. Thus Gunthorpe and Levy (1994) demonstrate that the planning horizon affects the portfolio composition; contrary to intuition, they find that short-term investors would invest in more aggressive assets, while long-term investors would invest in defensive assets. Thus they recommend that investors should first assess their transaction costs and then decide on their planning horizon. In models with no transaction costs, investors could rebalance their portfolio continuously without incurring any cost for those trades. In practice this is not the case, so Brown and Smith (2011) explored a simulation of three alternative heuristics in order to compare how strategies that disregard transaction costs compare with ones that do incorporate them. They find that strategies that rebalance portfolios periodically (one-period ahead, or many-months-ahead with monthly rebalancing) outperform a strategy that rebalances continuously. Moreover, the strategy that constructs portfolios for a longer period while allowing rebalancing more frequently than the planning horizon, performs slightly better than a strategy that employs a planning horizon equal to the rebalancing horizon. In our setting such a comparison would be between strategies that construct portfolios based on next-period idiosyncratic risk versus a strategy that constructs portfolios based on longer planning horizon (e.g. six months) even though portfolio rebalancing could still occur monthly; in this setting a portfolio construction based on a medium-term planning horizon would tend to outperform the strategy that constructs a portfolio using only next-month volatility, disregarding longer-term volatility. Importantly, as the planning horizon increases, volatilities quickly converge to the mean-reverting level, which could explain why investors would be using profitably the mean-reverting level instead of the next-period volatility. This is consistent with our exploratory analysis of the data, which showed that short-term forecasts (two-, three-month or longer) did not fare much better as explanatory variables than the one-month ahead forecasts, while for horizons over six months projected volatilities were numerically quite close to the mean-reverting level. Thus,

portfolios constructed using mean-reverting idiosyncratic volatilities could outperform portfolios based on one-month volatility forecasts even when portfolio rebalancing occurred on monthly frequency. Liu and Loewenstein (2002) explore the portfolio construction problem for CRRA investor in presence of transaction costs and finite horizon. They conclude that the optimal portfolio would be horizon-dependent, and would favour buy-and-hold strategy. Another perspective on the link between the portfolio planning horizon and the stockholding horizon is offered by Shiryayev et al. (2008), who employ a Black-Scholes model to analyse at what price level investors would be exiting optimally a given stock position. They demonstrate that such a choice would depend on the ratio between expected excess return and volatility, and for stocks for which that ratio is higher would be held longer; in case it exceeds 0.5, the optimal behaviour would be to hold the investment until the end of the planning horizon. Overall, those studies suggest that the prominence of the mean-reverting level is not just a data-mining artefact, but is consistent with rational portfolio construction based on intermediate planning horizons (e.g. six months or more) even though portfolios could still be rebalanced frequently; the reduced transaction costs from portfolio rebalancing could allow that approach to outperform myopic portfolios constructed on short-term forecasts of risk characteristics. Further theoretical arguments in that direction were recently proposed by Bichuch and Sircar (2015), who used perturbation methods to explore optimal investing in the presence of mean-reverting stochastic volatility and trading costs. In case of fast mean-reversion, they point out that there is just too little time for profit, and thus optimal trading decisions are based on the root mean squared volatilities, rather than on the current volatility factor realisation. On the other hand, when mean-reversion is slow, investors would not care for the average volatility, but rather for the Sharpe ratio. On average we observe fairly quick volatility reversion, which warrants the use of the expected mean-reverting level of volatility, rather than short-term expected volatility or averaged past values.

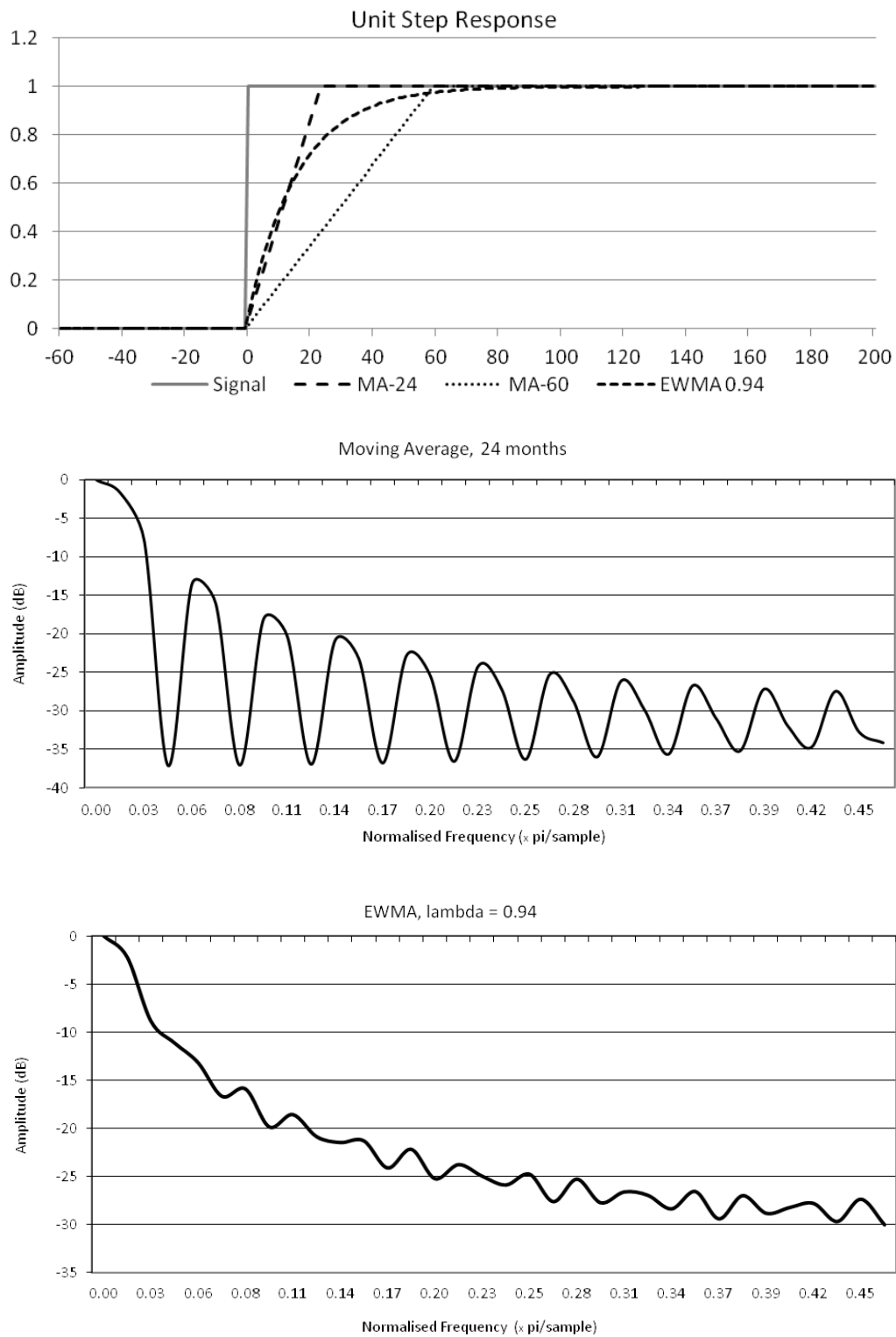
5.4. Comparison with other studies

Our approach highlights the consistency of existing evidence, rather than its divergence. We see that studies that found evidence supporting the studies of Merton and Levy used GARCH-based forecasts (Spiegel and Wang (2005); Fu (2009); Fu and Schutte (2010)) or filtered volatilities (Cao (2010); Cao and Xu (2010)), both of which would be correlated with the mean-reverting level. For example, in Table 11 on p. 133 we found that

forecasts from monthly GARCH(1,1) models correlated more with OLS residuals (0.92) than with Ang's historical volatility (0.65), which we found to be the second best predictor of true volatility. Likewise, its cross-sectional correlation with true ($IV_{i,t}^{egarch}$) volatility was just 0.79, similar to the results for OLS residuals (0.78) and the mean-reverting level (0.81), compared to a correlation to the ARMA forecasts of 0.91, which we found to be the best predictors of true volatility among the compared estimators. The point was made even more strongly in Table 13, where R^2 of \widehat{IV}_t^{garch} as predictor of true volatility was found to be less than half of the predictive performance of ARMA(1,1) forecasts and much closer to, albeit better than, OLS forecasts. Therefore, despite the forward-looking nature of the model per se, in practice GARCH forecasts with monthly data correlate better with OLS and the mean-reverting level, rather than future volatilities.

These findings should not be viewed as surprising. Indeed, the OLS variance forecast is the mean squared idiosyncratic shock in the respective month, and that squared shock is a measure of idiosyncratic variance in that month. Therefore, the OLS residuals serve as a moving-average filter of true volatilities, smoothing out transitory changes in a manner that is quite similar to the approach of Cao (2010) and Cao and Xu (2010).

Figure 4: Moving average and Exponentially-weighted moving average as filters



The relation between GARCH forecasts and the mean-reverting level might be more difficult to see immediately, but it is nonetheless there. Indeed, note that the EWMA specification ($\hat{\sigma}_t^2 = \lambda \hat{\sigma}_t^2 + (1 - \lambda)\epsilon_{t-1}^2$) is a particular case of the more general GARCH

model ($\hat{\sigma}_t^2 = \omega + \alpha \hat{\sigma}_t^2 + \beta \epsilon_{t-1}^2$). Moreover, the parameters we often observe in finance are sufficiently close to the assumption of EWMA – so much so, that they motivated the development of the Integrated GARCH model by NELSON, which builds on the observation that in many financial series $\alpha + \beta \approx 1$, as well as the incorporation of the EWMA in the RiskMetrics methodology (J.P.Morgan/Reuters (1996)). Thus, EWMA, like MA, could be viewed as an example of Finite Impulse Response (FIR) filter, where each value is calculated as a weighted sum of its most recent realisations:

$$y_n = b_0 x_{n-0} + b_1 x_{n-1} + b_2 x_{n-2} + \dots + b_N x_{n-N}.$$

In that setting, the moving average (the OLS volatility) is FIR filter with N between 24 and 60, where $b_i = \frac{1}{N}, i = 1, \dots, N$. Likewise, EWMA could be viewed as a FIR filter with $b_i = \frac{(1-\lambda)\lambda^{i-1}}{1-\lambda^N}, i = 1, \dots, N$. These filters do refer in terms of their response to unit step change of the impulse signal and amplitude reduction for difference frequencies (see Figure 4), however both OLS and EWMA are low-pass filters, i.e. they let low frequency changes through the filter while reducing high-frequency noise. The similarity between the GARCH(1,1) forecasts with SGED innovations and EWMA with SGED residuals (lambda estimated by best fit) is 0.94 in our sample, which confirms that GARCH(1,1) forecasts are very similar to those produced by EWMA and would thus filter primarily the low-frequency component of volatility. Therefore, we see that when one uses a forecasting method that functions as a low-pass filter, idiosyncratic volatility emerges as a significant predictor. This was the case with our tests involving OLS and GARCH(1,1), as well as in the related literature – the studies of Fu (2009), Fu and Schutte (2010) and Spiegel and Wang (2005) that used EGARCH(p,q), as well as the studies of Cao (2010) and Cao and Xu (2010), that used Hodrick-Prescott filters. Therefore, those studies produced significant correlation with returns, but the reason was not that they used accurate, forward-looking forecasts of volatilities, as suggested by Fu (2009), but rather filtered past volatilities to produce forecasts that correlated with the expected mean-reverting level of volatility.

Our study found furthermore that the best volatility forecasting method (among the ones compared in the study) was an insignificant predictor of the cross-section. We can hypothesise that the reason for its insignificance are the random changes from month to month, which are difficult to exploit economically due to transaction costs (Li et al. (2014)), and that results in longer planning horizons (Gunthorpe and Levy (1994)). However, that pattern could also give rise to other anomalies which resemble the ones documented in the related literature. In particular, let us decompose next-period expected volatility ($\hat{\sigma}_t$) into two

components: the mean-reverting level (m), and the transitory difference between mean-reverting volatility ($\hat{\sigma}_s = \hat{\sigma}_t - m$). Since $\hat{\sigma}_t$ is found to be an insignificant predictor of returns, while m is a significant one, then it follows algebraically that the coefficient $\hat{\sigma}_s$ would be negatively correlated with returns. Such considerations could explain the negative correlation between last-month volatility, documented by Ang et al. (2006), Ang et al. (2009), Li et al. (2014), and explored as a trading strategy by practitioners, e.g. Bender et al. (2013). However, since such studies use the last realised volatility, instead of the spread between expected volatility and the mean-reverting volatility, their results tend to be inconclusive. For example, the use of $\ln(IV_{t-1})$ as an explanatory variable found no evidence of significant negative correlation between volatility and returns. Moreover, the magnitude of the premium for idiosyncratic risk was found to be smaller in absolute terms compared to the premium for the mean-reverting level.¹⁵⁴ Therefore, our results suggest that contrary to the suggestion of Bender et al. (2013), if investors seek exposure to idiosyncratic risk, they should construct portfolios based on expected mean-reverting level of volatility, or, equivalently, the deviation of expected volatility from the mean-reverting level, as our study suggests that these are more reliable estimators of returns. Strategies that seek to profit from exposure to the extreme quantiles of the volatility distributions, like low-volatility or high-volatility portfolios, are likely to profit only inasmuch as those portfolios would be loaded with securities that have volatility significantly below the mean-reverting level (low-volatility investing), and volatility above the mean-reverting level. In the former case, the low-volatility portfolio would earn higher return not because there is some abnormal quality premium, but because those securities are priced based on their higher mean-reverting volatility. In the high-volatility portfolio the situation is reversed: for many of the securities the mean-reverting volatility is likely to be below the current one, and hence those portfolios would be earning lower returns than what is implied by their current volatility. Such considerations, however, do not exclude interaction with the mechanism documented by Fu (2009), namely the return reversals for some high-volatility stocks. Indeed, we documented that neither idiosyncratic volatility, nor most of the other explanatory variables explained the cross-section of returns in an environment of market turmoil. It could be conjectured, that idiosyncratic volatility may be insignificant in situations where there is great price uncertainty due to arrival of material new information to the market. Such considerations may be more relevant for high-volatility

¹⁵⁴ If we use the spread between expected volatility and the mean-reverting level, we would expect to see a slope that equals in absolute terms the one for the mean-reverting level, but with a negative sign.

portfolios, and therefore such securities may earn returns that deviate from their equilibrium expected returns. What we emphasise, however, is that the significance of the mean-reverting volatility and the insignificance of short-term forecasts is consistent with both positive correlation between returns and mean volatility, and with negative correlation between returns and the spread between expected volatility and the mean-reverting volatility.

The patterns documented in our study are also consistent with the observations of Bali and Cakici (2008) that the idiosyncratic volatility may not be a robust predictor of the cross-section. Such an outcome is consistent with tests that focus on larger stocks that have a wide investor base and a higher share of institutional ownership. Consistent with the underlying economic model, such securities would be earning lower premium for equivalent spreads of volatility, and thus the null hypothesis of no correlation would be more difficult to reject. Furthermore, as pointed out by Ang et al. (2010), tests based on portfolios may be less efficient, and may therefore suggest no correlation whereas in fact there is (failure to reject the null hypothesis when it is in fact false). We document similar patterns in our study: lower slopes for NYSE/Amex-traded stocks compared to Nasdaq-traded ones; less robust correlation when excluding stocks with lower unadjusted price compared to high-price stocks that have higher institutional ownership; less robust correlation for larger, presumably better known, stocks; less robust correlation when using portfolios and NYSE breakpoints. These findings did not contradict the underlying economic model *per se* – that would require widespread under-diversification among both individual and institutional investors, as well as across small and large-capitalisation stocks. Instead, investors seeking extra returns for idiosyncratic risk would likely gain more by focussing on less known stocks; such strategies may not benefit all investors, as those stocks may also have lower liquidity, and such trading strategies may have lower capacities. Nevertheless, studies that use models geared at predicting next-period volatility, which we demonstrated to be insignificant, would be also more likely to fail to establish a positive correlation between idiosyncratic risk and return simply because as they aim to predict the wrong volatility, the more accurately they predict volatilities, the less robust their findings would be.

5.5. Liquidity premium and other premia

Another difficulty in the empirical tests of the Merton model was that it predicted that premium was earned for exposure to idiosyncratic risk. Thus, researchers are expected to

demonstrate that idiosyncratic risk explains the cross-section of returns, and that idiosyncratic risk is not a proxy for any other explanatory variable. The former task is easier, as it requires proving a positive fact: that securities with higher idiosyncratic risk earn higher returns. The latter requires proving a negative fact: that there is no other (omitted) factor explaining the cross-section of returns. Even if the number of possible factors that could explain the cross-section was finite, many of them may not be available to researchers. The most obvious example would be Roll's critique: that the market portfolio that CAPM refers to includes all assets in the economy, rather than the exchange-traded shares (Roll, 1977). In that setting it could be argued that idiosyncratic risk served as a proxy of some other underlying risk. For example, Eiling (2013) suggested a link between idiosyncratic risk and industry-specific human capital. In our study we attempted to mitigate the problem of negative proof by using a wide range of controls. For example, we estimated idiosyncratic risk relative to the four-factor Fama–French–Carhart model in order to estimate residual variance after controlling as many factors as practically possible. We also employed a wide range of tests of robustness: a long list of control variables in the cross-sectional regressions (beta, size, book/market value, stock momentum, Roll's bid-ask spread, return reversals), as well as a range of subsamples (Nasdaq vs non-Nasdaq, unadjusted prices, growth vs recession episodes, market-wide volatility state, behavioural momentum, quartiles by correlated covariates); moreover, we also attempted to recover through statistical factor analysis any unobservable factor driving idiosyncratic returns that was omitted by the other tests.¹⁵⁵ Nevertheless, we recognise that there do remain other factors that could also be tested. Such factors could relate to the inadequate capture of risk preferences by expected return and standard deviation (i.e., first and second moments of the return process) and propose the use of further characteristics, e.g. skewness and kurtosis, and co-skewness and co-kurtosis with the market.¹⁵⁶ Relevant results on the link between idiosyncratic tail risk and the cross-section of returns (at daily frequency) were obtained by Huang, Liu, Rhee and Wu (2012), who reported that securities with higher extreme downside risk earned higher returns. In preparatory studies for this dissertation we have explored whether extreme downside risk could be useful in explaining the cross-section of returns, but we found the results inconclusive. Furthermore, it might be appropriate to explore the link to the extreme downside risk and the default probabilities, as extreme price changes may be

¹⁵⁵ See p. 206, Table 21: Fama–Macbeth cross-sectional regressions with loading on the principal factor affecting idiosyncratic returns, 07/1982 – 03/2013

¹⁵⁶ Kraus et al. (1976); Fang and Lai (1997); Harvey and Siddique (2000); Jondeau and Rockinger (2000); Dittmar (2002); Bali et al. (2011)

triggered by default events or revisions of risk of default. In principle such tests could be accomplished in various ways: external agency ratings, synthetic ratings or models akin to Altman's score¹⁵⁷, and option-based default probabilities¹⁵⁸. Due to lack of accounting data, we have not explored this avenue of research. We have nevertheless used high-betas and low prices as indications of higher probability of distress; these tests again confirmed the significance of idiosyncratic risk as an explanatory factor.¹⁵⁹

We observed significant correlation between the explanatory control variables in this study. Idiosyncratic risk was noted to correlate with size, beta and liquidity. The link between these four variables, however, may go well beyond simple correlation, so that it is possible that the true factor space is some lower-dimensional subspace of the size-beta-liquidity-volatility space. This issue could be particularly prominent in connection with stock liquidity. In this study we documented that the mean-reverting level of idiosyncratic volatility was a reasonably robust predictor of the cross-section of returns. This could not be said of liquidity. Indeed, in various specifications we observe changes in the magnitude and the sign of Roll's bid-ask spread. Of course, Roll's measure was just one of the possible measures of liquidity. The actual bid-ask spread, the measure of Amihud and Mendelson (1986), or the traded volume, were other possibilities. Due to lack of enough data we were unable to perform similar tests using actual bid-ask spreads and Amihud and Mendelson's bid-ask spreads. However, we did try to use the traded volume as a predictor and found that its coefficient was more often than not with the wrong sign in our sample, i.e. more traded stocks earned higher returns.¹⁶⁰ The unstable predictions yielded by the different facets of liquidity could be an argument against using liquidity as a predictor of returns. However, such conclusions could be unwarranted. The underlying assumption of the models of Levy and Merton was that investors were unable or unwilling to invest in some securities, e.g. due to transaction costs and finite divisibility of assets. However, there are also valid behavioural considerations in that decision to abstain from investing in certain assets. One of the principal investment constraints that should be documented in any individual investor's investment policy statement as summarised by the CFA Institute¹⁶¹ is the identification of any requirements for maintaining liquidity. For example, if there is uncertainty about an investor's

¹⁵⁷ Altman (1968)

¹⁵⁸ Merton (1974)

¹⁵⁹ See specification "k" in Table 18: Fama–Macbeth cross-sectional regressions with mean-reverting volatility – robustness checks, on p. 192.

¹⁶⁰ Such a finding was also reported by Malkiel and Xu (2004)

¹⁶¹ CFA Institute (2010)

income stream or risk of unplanned portfolio withdrawals, the resulting portfolio allocation should address these by increasing portfolio liquidity, e.g. increasing the share of fixed-income securities and exclusion of illiquid securities. Such considerations could be even more material for institutional investors, which may need to stand ready to convert their portfolios in cash to support any liquidity needs; for example, commercial banks are exposed to maturity mismatch risk as their deposits are callable at any time, while their assets are frozen in illiquid loans, a mismatch, the risk of which is mitigated by holding a portfolio of assets available for sale that can quickly be converted to cash in order to fund deposit withdrawals or outbound payments. Such considerations may result in investors excluding from their investment horizon securities with low liquidity. The liquidity preference could be incorporated directly in the portfolio-decision problem; one such approach was proposed by Lo et al. (2003), who used it in conjunction with five liquidity measures: the traded volume (total number of traded shares), natural logarithm of traded volume, turnover (traded volume/number of shares outstanding), percentage bid-ask spread, and Loeb price-impact function. Denoting the normalised liquidity metric by $l_{i,t}$, they suggested three alternative formulations of the portfolio-selection problem. The first approach is the liquidity-filtered portfolio, which is identical to the standard portfolio-selection problem but with the added constraint that only securities with normalised liquidity exceeding some threshold value (specific for each investor) are allowable investments. The important consequence is that the set of liquidity-filtered securities in this case corresponds directly to the set of securities the investor “knows” about in Merton’s model. Thus, the preference for liquidity in the formulation of Lo et al. (2003) would result in risk premium for idiosyncratic risk of the less liquid securities. This means that if all (or many) investors select their portfolio allocations based on liquidity filters, the premia for idiosyncratic risk would be just another name of the premium for liquidity; illiquid securities would earn higher return in equilibrium because the few investors in them are exposed to the undiversified idiosyncratic risk, while heavily-traded securities would earn no or low excess return for idiosyncratic risk because most of that risk is diversified. In this setting there is no real distinction between the liquidity premium and the idiosyncratic risk premium: the former simply refers to the underlying preference for liquidity, while the latter refers to the mechanism of how the aversion to illiquidity translates to risk premium through the channel of undiversified idiosyncratic risk. However, liquidity need not be the only filtering criterion: some investors may be excluding securities on other grounds, e.g. ethical (tobacco, weapons, etc.), transaction costs or creditworthiness (e.g. avoiding penny stocks), asset indivisibility (avoiding high-priced stocks), etc. All such

considerations could result in premia under the same mechanism as the premium for liquidity, which could be one reason why in this study idiosyncratic risk was found to be a more reliable predictor than liquidity. The two additional constraints proposed by Lo et al. (2003) could also have a similar impact on the equilibrium, although they do not map directly to the formulation of Merton. One such specification added a liquidity constraint to the problem, i.e. portfolio variance was minimised subject to achieving target levels of return and liquidity.¹⁶² Another specification included the portfolio liquidity as a penalty term in the quadratic optimisation problem. Again, these two formulations are closely related to the Markowitz formulation and should give rise to a similar blurring of the distinction between liquidity and idiosyncratic risk premia.

A similar pattern might be observable also in terms of capitalisation constraints: small companies might be disliked by investors on various grounds, e.g. they may have more risky (less diversified) cash flows or less mature corporate governance. Neither of these actually guarantees superior performance¹⁶³, but if investors avoid small companies (even when they are liquid¹⁶⁴) and filter them out when making portfolio allocations, this again might give rise to idiosyncratic risk premium accruing to small stocks. Similarly, larger investors may dislike small stocks simply because their free float is too small for the typical capacities of their trading strategies, i.e. these investors do not engage in too small transactions for efficiency reasons, which again could result in a risk premium for small stocks, as documented by Fama and French (1992).

For our cross-sectional tests such considerations could have material impacts. If the premia for size, liquidity and idiosyncratic risk overlap, then the cross-sectional regressions may suffer from multicollinearity. When such multicollinearity is not addressed, the OLS coefficients are still unbiased, but their standard errors are higher, which may explain the lack of robust performance of the liquidity and size variables in the cross-sectional tests. For empirical applications, however, addressing this problem explicitly may result in superior

¹⁶² The liquidity constraint in that case has the following form:

$$l_0 = \begin{cases} w' l, & \text{for long-only portfolio,} \\ \sum_{i=1}^n \frac{|w_{i,p}|}{\sum_{j=1}^n |w_{j,p}|} l_{i,i}, & \text{for long-short portfolio.} \end{cases}$$

¹⁶³ See Ch.6 in Damodaran (2004)

¹⁶⁴ In principle the turnover usually refers to the ratio of the number of shares traded over the number of shares outstanding. There may be heavy trading in the shares of some small company, yet the value of the traded volume may be relatively small and uninteresting for larger, institutional investors because of the limited capacity of the position.

forecasts and more stable portfolio allocations.

5.6. Is it tradable?

The significance of idiosyncratic volatility as a characteristic explaining the cross-section of returns was confirmed by comparing the implied premia for one standard deviation difference of each characteristic. In the case of idiosyncratic volatilities (Table 16 on p. 160), our results imply an annualised premium for idiosyncratic risk of between 2.58 (in specification 3) and 3.17 (specification 6), which was well above the premia for beta (0.86), capitalisation (1.58), and liquidity (0.76).¹⁶⁵ This finding was consistent with similar results obtained by Fu (2009) and Spiegel and Wang (2005), albeit those studies used different methodology to predict idiosyncratic volatility.

The implied premium for idiosyncratic risk was found to be statistically significant. However, its economic significance was somewhat less clear. A premium of between 2.58 and 3.17 seemed to merit consideration. On the other hand, if the expected mean-reverting levels of volatility changed significantly over time, then the trading costs of executing a strategy that was betting on idiosyncratic risk premium could offset most or the entire idiosyncratic risk premium. There are two types of costs relevant for deal execution: explicit costs (mostly brokerage fees) and implicit costs (bid-ask spreads, execution costs). Brokerage fees are incurred due to the remuneration of the intermediary executing each trade. The bid-ask spread is the difference between sell and buy prices, while the execution costs measure the actual price, at which an order of a given price could be executed. Another type of implicit cost is the opportunity cost that refers to the foregone profit due to partial order execution or cut-back of the intended transaction volume. There is a significant body of literature concerning the measurement of the implicit and explicit costs. The review article by Keim and Madhavan (1998) summarises the key findings of those studies. Concerning brokerage fees, Keim and Madhavan (1998) report that these costs change over time, generally in downwards direction, and average about 0.20% of the traded volume. The broker fees depend on the price of the traded stock, as well as on broker type, trading mechanism, order difficulty, type of trade, and exchange. For example, ECN crossing trades¹⁶⁶ are

¹⁶⁵ Based on specification #6.

¹⁶⁶ “Electronic Communications Networks, or ECNs, as defined in Rule 600(b)(23) of

reported to incur a cost of 1-2 cents per share, whereas upstairs broker-dealers¹⁶⁷ could charge up to 10-15 cents per share. Keim and Madhavan (1998) also point out that brokerage fees do not take into account the additional soft-dollar services rendered by the brokers to their clients, and as such may overestimate the true economic price of brokerage services.

The quoted bid-ask spread measures the costs to execute a buy or sell transaction whereupon the buyer would need to offer the ask price or the seller would need to lower the ask price to the bid price if a sale is to occur. As such, the spread measures the overhead that the buyer or seller would incur in order to execute a transaction. The reviewed literature found that the quoted bid-ask spread depended strongly on the company size and liquidity, with the bid-ask spread as low as 0.5% for the larger-capitalisation liquid stocks, and up to 4-6% for small and illiquid stocks. Keim and Madhavan (1998), however, also point out three considerations why the quoted bid-ask could be overstating the true cost incurred by the traders. Firstly, they point out that trades often occur inside the quoted bid-ask spread. They also note that bid and ask prices have a systematic tendency to follow the market, with both bid and ask prices increasing following a buy order and decreasing following a sale order. Finally, they point out that large block transactions need not close at the bid or ask prices. We should also recognise, that the context of the study of Keim and Madhavan (1998) is institutional trades, and the second point (that quotes follow market direction) may be more relevant for higher-frequency trading, but less so in strategies seeking exposure to idiosyncratic risk. Likewise, idiosyncratic premium should be higher for less-known stock, where larger block trades are less likely to occur.

Another consideration of the cost analysis of trades concerned the execution costs. These refer to the limited depth of the order book, so that not the entire orders could be

Regulation NMS, are electronic trading systems that automatically match buy and sell orders at specified prices. ECNs register with the SEC as broker-dealers and are subject to Regulation ATS. Subscribers, which are typically institutional investors, broker-dealers, and market-makers — can place trades directly with an ECN. Individual investors must currently have an account with a broker-dealer subscriber before their orders can be routed to an ECN for execution. When seeking to buy or sell securities, ECN subscribers typically use limit orders. ECNs post orders on their systems for other subscribers to view. The ECN will then automatically match orders for execution. An ECN may choose to facilitate compliance by a market-maker with its obligations under the Commission's Quote Rule by transmitting the ECN's best bid/offer to a national securities exchange or registered securities association for public display.” (SEC (n.d.))

¹⁶⁷ The Nasdaq financial glossary defines upstairs market as “A network of trading desks for the major brokerage firms and institutional investors, which communicate with each other by means of electronic display systems and telephones to facilitate block trades and program trades.” (Nasdaq, n.d.)

executed at the bid or ask price. For example, if a seller wished to dispose of a larger block of shares at market prices, the best bid order would be executed, after which the seller would need to go to the next available buy-order, and then to the next until the entire block is sold. Alternatively, the sale could be split into smaller sales so that the transaction can have lower impact on market prices. Therefore, a strategy that trades less liquid shares would also have higher execution costs and/or limited capacity, i.e. smaller volume that could be invested in that strategy.

The aforementioned arguments suggest that even if the idiosyncratic risk premium exists, the opportunities for its use in trading might be limited as costs could offset the gains from higher exposure to idiosyncratic risk. Keim and Madhavan (1998) estimate the total costs of equity trades of a set of institutional investor trades broken down by type of trade (buyer-initiated vs seller-initiated), market (NYSE or Amex exchanges vs Nasdaq), trade size quartiles and capitalisation quintiles. They find that trading costs increase steeply with the increase of order size. Thus, the lowest quartile exchange-listed buyer-initiated trades incur a cost of 0.31% compared to 0.90% for the highest-quartile trades; the corresponding numbers for Nasdaq-listed stocks are about twice those numbers – 0.76% and 1.80%, respectively. Thus, increasing the size of the trades also increases the total costs and reduces the gains from the trading strategy. In Table 18 on page 170 we reported that Nasdaq shares earned a two to three times higher premium compared to NYSE/Amex-traded stocks. Thus, the higher gains on exposing to idiosyncratic risk on Nasdaq could be largely offset by the higher trading costs. Whether that is indeed the case requires further research. Indeed, the argument concerning the higher costs of larger trades is generic for any trading strategy. Furthermore, the split reported in Keim and Madhavan (1998) is different from the split used in Table 18; the former is based on exchange listing, while the Datastream assignment is based on exchange where most of the trade occurred, thus limiting the comparability between the premia and costs.

In a similar vein, Keim and Madhavan (1998) also report that the total traded costs decrease with the market capitalisation. Thus, trading costs for buyer-initiated trades in exchange-traded stock in the highest-capitalisation quintile amount to 0.31% and increased non-linearly to 1.78% in the smallest-quintile stocks. The corresponding total costs for Nasdaq-listed stocks amounted to 0.24% and 2.85%, respectively, confirming the generally higher trading costs for Nasdaq-listed shares.

Overall, we find that idiosyncratic risk is a significant factor in explaining the cross-section of returns. The tradability of that finding, however, is yet to be confirmed. The

recent study of McLean and Pontiff (2016) may be particularly relevant in that matter, as it uncovers that markets rapidly exploit abnormal returns, albeit not to the full extent. Since the return to idiosyncratic volatility is an equilibrium phenomenon rather than an anomaly, we should expect that it should continue to be a reliable predictor of the cross-section. However, we point out that the multicollinearity in the cross-sectional regressions may be a both theoretically and empirically sound way to carry out such tests, and portfolio sorts should first aim to estimate the true dimension of the size-liquidity-volatility space.

6. Conclusions and directions for future research

6.1. Conclusions

This study explored whether idiosyncratic volatility was associated with higher equilibrium returns, as predicted by the models of Levy (1978), Merton (1987), and Malkiel and Xu (2004). The existing empirical studies provided mixed support to those CAPM extensions. Thus, the studies of Ang et al. (2006) and Ang et al. (2009) found significant negative correlation between idiosyncratic risk and returns. In contrast, the studies of Malkiel and Xu (2004), Fu (2009) and Spiegel and Wang (2005) documented significant positive correlation, while the tests of Bali and Cakici (2008) found no significant correlation. The studies of Cao (2010) and Cao and Xu (2010) proposed that it is the long-term component of idiosyncratic volatility that explains the cross-section of returns. Most of those studies employed different measures of idiosyncratic risk, which hindered the interpretation of those conflicting findings. Some of these tests were backward-looking (e.g. Ang et al., 2006, or Cao, 2010), whereas Fu (2009) pointed out that investors would base their portfolio decisions on expected future values, rather than on just historical ones. On the other hand, the results of Fu (2009) were questioned for omitting last-month return (omitted variable bias) or for inclusion of current-month return in generating expected volatilities (look-ahead bias).

Against this backdrop we compared the predictive performance of the main classes of estimators of idiosyncratic volatility, as well as their performance in predicting future returns. We used four estimators of returns: the historical volatility employed by Ang et al. (2006); the forward-looking ARMA(1,1); the GARCH(1,1); and the OLS estimator. We found that the latter two significantly underperformed in forecasting next-period realised volatility, but in contrast confirmed the positive link between idiosyncratic risk and returns. The contrast between poor performance in forecasting volatilities and good performance in predicting returns demonstrated that contrary to expectations, the results of Fu (2009), Spiegel and Wang (2005), and Cao (2010) were not an outcome of superior forecasting performance; in fact, the measure of Ang et al. (2006) proved to be a superior predictor of the cross-section compared to GARCH forecasts with monthly data. Moreover, we found that the negative correlation reported by Ang et al. (2006) is an artefact of poor performance of very high-volatility (distressed) stocks, and after we reduced the skewness of the cross-sectional distribution of idiosyncratic volatilities, the estimator of Ang et al. (2006) became insignificant.

We found that idiosyncratic risk and stock returns were nevertheless robustly correlated, but returns correlated with the mean-reverting level of volatility, rather than next-month volatility. It proved to be a very robust predictor of the cross-section of stock returns that remained statistically and economically significant after controlling for beta, size, book/market, momentum, return reversals, liquidity, portfolio construction, change of frequency (from monthly to daily data), omitted factors, unadjusted price, credit risk, and for primary exchange, on which shares were traded. We found that the strength of the correlation between mean-reverting volatility and returns was not constant, but depended on the primary market, on which the stock was traded, with a regression slope for Nasdaq-traded stocks almost three times that of NYSE and Amex stocks. This finding is consistent with observations that NYSE and Amex are preferred by larger investors, while stocks traded at Nasdaq are preferred by smaller and presumably – less diversified investors. Thus, differences in slopes between Nasdaq and NYSE-traded stocks are fully-consistent with the underlying economic models that predict that the premium per unit of idiosyncratic risk would be lower the better known the stock.

We also documented that the link between idiosyncratic risk and return does not hold in high-volatility environments. Moreover, in periods of recession the premium for idiosyncratic risk is significantly higher compared to expansion periods, and while the direction of that difference is not surprising, its magnitude poses a puzzle.

Our findings are significant from both theoretical and practical perspective. From theoretical perspective they support the predictions of the models of Levy (1978) and Merton (1987), and suggest that under-diversification is sufficiently wide-spread to result in existence of a material risk premium. Moreover, our results suggest that investors avoid frequent rebalancing of their portfolios, probably due to the associated transaction costs, and base their investment decisions not on short-term horizons, but on medium-term characteristics. The observed material speed of mean-reversion, consistent with previous results of stationarity of volatilities, suggested that volatilities normally converged to the mean-reverting level in less than six months, so this might be a suitable target investment horizon for many investors. This would be particularly true for non-institutional investors, which are likely to trade in better-known stocks due to market depth considerations, a segment where the correlation between idiosyncratic risk and returns is low. In the interpretation of our results we particularly emphasised the process of portfolio construction with additional constraints, especially on liquidity, and we pointed out that such a portfolio construction framework would blur or even remove the difference between liquidity, size and idiosyncratic risk

premia. Therefore, our results suggest that portfolio planning and horizon deserve further exploration and the assumption of continuous rebalancing may not be a realistic one for all segments of the market.

For practitioners our results should be useful as we identify a characteristic that predicts the cross-section of returns. Unlike other studies in the field, this predictive performance is implied by the theory and does not constitute a stock market anomaly; it should therefore persist in time. Failure to incorporate that characteristic in portfolio performance metrics could result in a false sense of gaining alpha whereas the portfolio could be simply exposed to idiosyncratic risk. Similarly, constructing portfolios aimed to exploit the negative correlations documented by Ang et al. (2006) could expose investors to stocks with very high volatility. On the other hand, portfolios constructed based on the mean-reverting volatilities could result in superior outcomes for investors due to persistence of those volatilities that requires less frequent portfolio adjustments. Nevertheless, such portfolios should also balance idiosyncratic volatilities with exposure to lower liquidity and shallow markets. Our analysis of the results also suggests that idiosyncratic risk is closely intertwined with liquidity risk and other characteristics that may be disliked by investors, e.g. small size. Therefore, the hedging of risk exposures might be better achieved by explicitly accounting for the dependence between various characteristics (e.g. high idiosyncratic volatility being associated with higher beta, lower liquidity, smaller size, higher default risk).

In our view, future research should seek to improve the forecasting of the mean-reverting level of volatility by directly using higher-frequency data. Furthermore, dimensionality reduction could be employed in order to identify the common factors driving stock returns. Such dimension reduction should be useful for both theoretical reasons (identification of factors based on theory, rather than data mining), as well as practical considerations. If liquidity preference results in avoidance of smaller, less liquid stocks, giving rise to idiosyncratic risk premium, then the true premium to the exposure of the underlying factor would be underestimated by using many characteristics loaded on that factor and trying to construct portfolios on each separate characteristic.

6.2. Limitations and directions for future research

We explored the link between idiosyncratic volatility and returns using the methodology of Fama and MacBeth (1973) with individual securities as assets and Newey

and West (1987) standard errors. The existence of that link was predicted by the models of Levy (1978) and Merton (1987). These models predicted that if for some reason investors were unable to diversify idiosyncratic risk fully, then they would be requiring compensation for the undiversified idiosyncratic risk. However, since we do not observe the actual portfolios held by all market participants, it was not possible to estimate reliably the undiversified component of idiosyncratic risk. One of the few studies that actually attempt to proxy undiversified idiosyncratic risk was that of Malkiel and Xu (2004). However, their approach did not measure the undiversified risk but rather used idiosyncratic volatility of quantile portfolios in place of individual security volatility, arguing that the former could be closer to undiversified volatility than the latter. Nonetheless, it was not clear why investors should be diversifying their portfolios using similar securities from the same quantile bucket, nor whether the number of securities in each bucket was consistent with portfolio sizes observed by some empirical studies; for example, if a bucket contained 50 securities, while investors held 10 securities, then the undiversified risk could be understated. Such underestimation of the true value and the associated reduction of the cross-sectional differences in the explanatory variable could result in false negatives in the cross-sectional tests.¹⁶⁸ Similar criticisms could be raised against the second point: greater diversification would be obtained by investing in securities with lower correlation, which are likely to be in another size-beta portfolio than the tested one, which could result in overstating the undiversified idiosyncratic risk. Based on the lack of normative grounds to choose that approach, in this study we used almost exclusively the total idiosyncratic risk as a predictor variable. Nevertheless, we have also studied portfolios as assets and we have found that portfolio alphas increase with the increase of volatilities, which addresses concerns that the obtained results are driven by data errors.¹⁶⁹ Most other studies, including ours, used total idiosyncratic risk of individual securities. The problem with using total idiosyncratic risk was that the share of the undiversified component in total idiosyncratic risk was not constant across securities, and to a presumably lesser extent – in time. Thus, securities that were widely followed (present in many portfolios) should *ceteris paribus* have lower undiversified idiosyncratic risk. Since we were unable to observe the actual portfolios held, it was not possible to observe the risk-return frontier available to each investor and deduce its impact on equilibrium prices and returns. Therefore, studies that incorporate some information on the

¹⁶⁸ cf the arguments in Fama and French (1992); Ang et al. (2010)

¹⁶⁹ See page 218, Table 25: Portfolio Alphas Relative to Fama–French–Carhart Model.

breadth of the investor base provide valuable evidence on the consistency of the correlation between idiosyncratic risk and returns.

A significant difference across the related studies concerned the measurement of idiosyncratic risk. In the first place it should be emphasized that idiosyncratic volatility is just one of the possible measures of idiosyncratic risk. It is indeed the one most commonly used, but its choice may be dictated by convenience and concerns about generalizability (external validity) of the research; in particular, as discussed previously, that was one of our primary considerations in selecting our research approach. For example, Ang et al. (2006) and Brockman et al. (2009) used dispersion of analysts' forecasts as measures of idiosyncratic risk, albeit in the context of control variables. Unfortunately, the availability of such alternative proxies of idiosyncratic risk was limited, which constrained the available sample both in terms of number of followed securities and past information. On the other hand, idiosyncratic volatility had the advantage of being a function of past prices, which were available for all traded securities. Moreover, there existed established models for volatility forecasting like GARCH. Furthermore, volatilities were an established measure of risk since the formulation of the mean-variance optimisation problem, as formulated by Markowitz (1952). Nevertheless, inasmuch as return distribution need not be specified fully in terms of its mean and variance, other measures of idiosyncratic risk could also yield further useful insights. For example, Huang, Liu, Rhee and Wu (2012) found that extreme downside risk also predicted the expected returns. Other possible idiosyncratic risk measures could include the third and fourth moments (skewness and kurtosis), and their dependence with the market factor (co-skewness and co-kurtosis).

Behavioural patterns might be a particularly promising direction for future research. For example, Kraus et al. (1976) suggested on classical grounds that investors should prefer co-skewness with the market, while the behavioural model of Barberis and Huang (2008) identified preference for asset's own skewness. Iwasawa and Uchiyama (2013) similarly suggested that institutional investors prefer high-beta securities in an attempt to beat the market, while individual investors like stocks with high tail return (positive skewness) which they call "gambling preference". We noted a significant change of slope of idiosyncratic risk during economic expansion and downturn – that split was, in fact, resulting in a larger spread between the slopes than any other split, including those based on liquidity or stock exchange (*cf.* Table 18 on p. 170). The slope during contraction periods was materially higher compared to expansion periods. We pointed out that changes in risk tolerance could hardly explain such a differential between the premia. Furthermore, the idiosyncratic risk premium is driven by

under-diversification, a problem that is more relevant for individual and smaller investors. However, such changes in risk tolerance could be compounded with some cyclical, flight-to-safety behaviour of investors. For example, Papaioannou et al. (2013) identify five major factors driving pro-cyclical investment behaviour, *viz.*: (i) underestimation of liquidity needs in downturn environment; (ii) uncertainty in assessing market risk, compounded with decreased transparency of issuers; (iii) incentive problems, e.g. due to focus on short-term performance, resulting in increased investment in illiquid and more speculative assets; (iv) reporting and disclosure requirements concerning loss-making positions and portfolios; and (v) accounting and regulatory requirements like strict mark-to-market policies resulting in immediate loss recognition and eating up capital of institutional investors, as well as herding, e.g. due to common, market-standard valuation models. Such considerations might plausibly result in flight to stocks with lower volatility and higher capitalisation (liquidity), and could explain the significant increase in the slope of idiosyncratic risk in the cross-sectional regressions during recessions.

In this study idiosyncratic risk was measured as the standard deviation of returns that is not explained by changes of economic factors like market excess return, Fama and French's small-minus-big and high-minus-low factors, and the momentum factor. Security returns were measured by monthly arithmetic excess returns. We acknowledge that there is not a single correct way of measuring either idiosyncratic return, or idiosyncratic volatility. Thus, idiosyncratic returns could be measured relative to the CAPM model, or relative to the Fama–French factor model, or relative to Fama–French–Carhart model, as we did. However, nothing prevents the use of any other model of asset returns, for example, a macroeconomic model or a stochastic factor model. Malkiel and Xu (2004) pointed out that different models were unlikely to result in very different conclusions, as idiosyncratic volatility was a second moment. We agree with that point, but we have also managed to mitigate that risk by using a factor model with four explanatory variables and allowing for some unidentified factor recovered through heteroscedastic factor analysis. If such a significant contributory factor had been omitted from the Fama–French specification, that would have resulted in correlation of idiosyncratic returns and would have been captured by the statistical factor. Whenever possible we opted for the Fama–French–Carhart model in order to mitigate possible concerns of the presence of an omitted momentum factor. Moreover, in Section “4.5.1. Was there an omitted factor?” we additionally tested whether some omitted factor could account for the significance of idiosyncratic volatility in explaining the cross-section of returns and we found that that was not the case. Nonetheless, one cannot rule out entirely that some differences in

our findings from previous studies could be the result of our choice of factor model. However, we point out that the use of a more comprehensive model should be reducing the significance of idiosyncratic volatility as an explanatory factor, rather than increasing it. Therefore, we are sceptical that our results could be driven by the choice of a factor model, and we expect that a study using a lower number of factors would result in qualitatively similar results.

Idiosyncratic risk should explain returns in equilibrium. In Table 18¹⁷⁰ on p. 170 we explored how cross-sectional regressions were affected in the three volatility regimes and found that the results in the high-volatility regime deviated from the theory. Part of the problem could be the low number of months in that regime, and correspondingly the difficulty to obtain more accurate estimates of the regression coefficients and the standard errors, thus resulting in the inconclusive results. However, the high-volatility episodes may reflect adjustment of investor expectations and risk tolerance, or macroeconomic shocks. The standard cross-sectional models do not include adjustment for such common shocks and that may result in higher standard errors of the estimates and failure to reject the null hypothesis in the full sample. Such considerations may help explain why the mean-reverting level of volatility was found to be a significant predictor of the cross-section of returns, but one-step forecasts produced insignificant results. For example, Andrews (2005) explored the impact of common shocks in cross-sectional regressions and found that the resulting coefficient estimates were consistent as long as the errors were uncorrelated with the regressors conditional on the sigma field generated by the common shocks. In that case he proved that the t -, F - and Wald tests were also asymptotically valid. However, if the errors were correlated with the regressors conditional on the sigma-field generated by the shocks, then the null rejection probabilities for t -, F - and Wald tests converged to one. This might be a relevant consideration in the empirical tests of idiosyncratic volatilities, as some common shocks (e.g. economic shocks or technological trends before the dot-com bubble) might have resulted in dependence between errors and some of the explanatory variables like liquidity and idiosyncratic risk.

¹⁷⁰ Table 18: Fama–Macbeth cross-sectional regressions with mean-reverting volatility – robustness checks

References

- Allais, M. (1953) 'Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine.' *Econometrica*, 21(4) pp. 503–546.
- Altman, E. I. (1968) 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy.' *Journal of Finance*, 23(4) pp. 189–209.
- Amihud, Y. and Mendelson, H. (1986) 'Asset pricing and the bid-ask spread.' *Journal of Financial Economics*, 17(2) pp. 223–249.
- Andersen, T. G. and Bollerslev, T. (1998) 'Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts.' *International Economic Review*, 36(4) pp. 885–905.
- Andrews, D. W. K. (2005) 'Cross-Section Regression with Common Shocks.' *Econometrica*, 73(5) pp. 1551–1585.
- Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006) 'The Cross-Section of Volatility and Expected Returns.' *Journal of Finance*, 61(1) pp. 259–299.
- Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2009) 'High idiosyncratic risk and low returns: International and further US evidence.' *Journal of Financial Economics*, 91 pp. 1–23.
- Ang, A., Liu, J. and Schwarz, K. (2010) *Using Stocks or Portfolios in Tests of Factor Models. American Finance Association 2009 Meetings Paper.*
- Aronson, D. (2007) *Evidence-Based Technical Analysis*. New Jersey: John Wiley & Sons, Inc.
- Arrow, K. J. (1984) *Collected Papers of Kenneth J. Arrow, Volume 3: Individual Choice under Certainty and Uncertainty*. The Belknap Press of Harvard University Press.
- Asparouhova, E., Bossaerts, P., Roy, N. and Zame, W. (2016) "Lucas" in the Laboratory.' *Journal of Finance*, 71(6) pp. 2727–2780.
- Bai, J. and Ng, S. (2002) 'Determining the Number of Factors in Approximate Factor Models.' *Econometrica*, 70(1) pp. 191–221.
- Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996) 'Fractionally integrated generalized autoregressive conditional heteroskedasticity.' *Journal of Econometrics*, 74 pp. 3–30.
- Bali, T. G. and Cakici, N. (2008) 'Idiosyncratic Volatility and the Cross Section of Expected Returns.' *Journal of Financial and Quantitative Analysis*, 43(1) pp. 29–58.
- Bali, T. G., Cakici, N. and Whitelaw, R. F. (2011) 'Maxing out: Stocks as lotteries and the cross-section of expected returns.' *Journal of Financial Economics*, 99(2) pp. 427–446.
- Barberis, N. and Huang, M. (2008) 'Stocks as Lotteries: The Implications of Probability

- Weighting for Security Prices.’ *American Economic Review*, 98(5) pp. 2066–2100.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008) ‘Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise.’ *Econometrica*, 76(6) pp. 1481–1536.
- Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970) ‘A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains.’ *Annals of Mathematical Statistics*, 41(1) pp. 164–171.
- Benartzi, S. and Thaler, R. H. (1995) ‘Myopic Loss Aversion and the Equity Premium Puzzle.’ *Quarterly Journal of Economics*, 110(1) pp. 73–92.
- Bender, J., Briand, R., Melas, D. and Subramanian, R. A. (2013) *Foundations of Factor Investing*. (Research Insight).
- Bhootra, A. and Hur, J. (2014) ‘High Idiosyncratic Volatility and Low Returns: A Prospect Theory Explanation.’ *Financial Management (forthcoming)*.
- Bichuch, M. and Sircar, R. (2015) *Optimal Investment with Transaction Costs and Stochastic Volatility Part I: Infinite Horizon*.
- Black, F. (1993) ‘Beta and Return.’ *Journal of Portfolio Management*, 20(1) pp. 8–18.
- Black, F., Jensen, M. C. and Scholes, M. (1972) ‘The Capital Asset Pricing Model: Some Empirical Tests.’ In: Praeger Publishers Inc.
- Blume, M. E. and Friend, I. (1975) ‘The Asset Structure of Individual Portfolios and Some Implications for Utility Functions.’ *Journal of Finance*, 30(2) pp. 585–603.
- Bollerslev, T. (1986) ‘Generalized Autoregressive Conditional Heteroscedasticity.’ *Journal of Econometrics*, 31 pp. 307–327.
- Bondt, W. F. M. De and Thaler, R. (1985) ‘Does the Stock Market Overreact?’ *Journal of Finance*, 40(3).
- Brandt, M. W., Brav, A., Graham, J. R. and Kumar, A. (2010) ‘The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes?’ *Review of Financial Studies*, 23(2) pp. 863–899.
- Brockman, P., Schutte, M. G. and Yu, W. (2009) ‘Is Idiosyncratic Risk Priced? The International Evidence.’ *Manuscript available at SSRN*, July.
- Brooks, C., Li, X. and Miffre, J. (2011) ‘Idiosyncratic Risk and the Pricing of Poorly-Diversified Portfolios.’ *EDHEC-Risk Institute Working Papers*, May.
- Brown, D. B. and Smith, J. E. (2011) ‘Dynamic Portfolio Optimization with Transaction Costs: Heuristics and Dual Bounds.’ *Management Science*, 57(10) pp. 1752–1770.
- Caldwell, B. (1980) ‘Positivist Philosophy of Science and the Methodology of Economics.’ *Journal of Economic Issues*, XIV(1) pp. 53–76.

- Campbell, J. Y., Lettau, M., Malkiel, B. G. and Xu, Y. (2000) 'Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk.' *National Bureau of Economic Research Working Papers*, 7590, March.
- Cao, X. (2010) *Three essays on the puzzles in finance*. The University of Texas at Dallas.
- Cao, X. and Xu, Y. (2010) 'Long-Run Idiosyncratic Volatilities and Cross-Sectional Stock Returns.' *Available at SSRN*, February.
- Carhart, M. M. (1997) 'On Persistence of Mutual Fund Performance.' *Journal of Finance*, 52(1) pp. 57–82.
- CFA Institute (2010) *Elements of an Investment Policy Statement for Individual Investors*.
- Chamberlain, G. (1983) 'A characterization of the distributions that imply mean-variance utility functions.' *Journal of Economic Theory*, 29(1) pp. 185–201.
- Chan, L. K. C. and Lakonishok, J. (1992) 'Robust Measurement of Beta Risk.' *Journal of Financial and Quantitative Analysis*, 27(2) pp. 2265–2282.
- Chen, N.-F., Roll, R. and Ross, S. A. (1986) 'Economic Forces and the Stock Market.' *Journal of Business*, 59(3) pp. 383–403.
- Chicheportiche, R. and Bouchaud, J.-P. (2012) 'The Joint Distribution of Stock Returns in not Elliptical.' *International Journal of Theoretical and Applied Finance*, 15(3) pp. 1–23.
- Cochrane, J. H. (2005) *Asset Pricing*. Revised Ed, Princeton University Press.
- Connor, G. (1995) 'The Three Types of Factor Models: A Comparison of Their Explanatory Power.' *Financial Analysts Journal*, 51(3) pp. 42–46.
- Connor, G., Goldberg, L. R. and Korajczyk, R. A. (2010) *Portfolio Risk Analysis*. Princeton University Press.
- Connor, G. and Korajczyk, R. A. (1993) 'A Test for the Number of Factors in an Approximate Factor Model.' *Journal of Finance*, 48 pp. 1263–1291.
- Cooper, M. J., Gulen, H. and Schill, M. J. (2008) 'Asset Growth and the Cross-Section of Stock Returns.' *Journal of Finance*, LXIII(4) pp. 1609–1651.
- Damodaran, A. (2004) *Investment Fables: Exposing the Myths of 'Can't Miss' Investment Strategies*. Financial Times Prentice Hall.
- Daniel, K. and Titman, S. (1997) 'Evidence on the Characteristics of Cross Sectional Variation in Stock Returns.' *Journal of Finance*, 52(1) pp. 1–33.
- Daniel, K. and Titman, S. (1998) 'Characteristics or Covariances?' *The Journal of Portfolio Management*, 24(4) pp. 24–33.
- Davis, M. H. A. (2004) 'Complete-Market Models of Stochastic Volatility.' *In Proceedings: Mathematical, Physical and Engineering Sciences*. The Royal Society, pp. 11–26.

- DeGenarro, R. P. and Robotti, C. (2007) *Financial Market Frictions. Federal Reserve Bank of Atlanta Economic Review.*
- Dickey, D. A. and Fuller, W. A. (1979) 'Distribution of the Estimators for Autoregressive Time Series With a Unit Root.' *Journal of the American Statistical Association.* Taylor & Francis, Ltd. on behalf of the American Statistical Association, 74(366) pp. pp. 427–431.
- Diebold, F. X. and Mariano, R. S. (1995) 'Comparing Predictive Accuracy.' *Journal of Business and Economic Statistics*, 13 pp. 253–265.
- Dittmar, R. F. (2002) 'Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns.' *Journal of Finance*, 57(1) pp. 369–403.
- Eiling, E. (2013) 'Industry-Specific Human Capital, Idiosyncratic Risk, and the Cross-Section of Expected Stock Returns.' *Journal of Finance*, 68(1) pp. 43–84.
- Elton, E. J. and Gruber, M. J. (1977) 'Risk Reduction and Portfolio Size: An Analytical Solution.' *Journal of Business*, 50(4) pp. 415–437.
- Embrechts, P., Kluppelberg, C., Mikosch, T., Kuppelberg, C. and Mikosch, T. (2003) *Modelling Extremal Events for Insurance and Finance.* Springer Verlag Berlin.
- Engle, R. F. (1982) 'Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation.' *Econometrica*, 50 pp. 987–1008.
- Engle, R. F. and Patton, A. J. (2001) 'What Good is a Volatility Model?' *Quantitative Finance.* (Research Paper), 1 pp. 237–245.
- Fama, E. F. and French, K. R. (1992) 'The cross-section of expected stock returns.' *Journal of Finance*, 47(2) pp. 427–465.
- Fama, E. F. and French, K. R. (1993) 'Common risk factors in the returns of stocks and bonds.' *Journal of Financial Economics*, 33, March, pp. 3–56.
- Fama, E. F. and French, K. R. (1996) 'Multifactor explanations of asset pricing anomalies.' *Journal of Finance*, 60(1) pp. 55–84.
- Fama, E. F. and French, K. R. (2007) 'The Anatomy of Value and Growth Stock Returns.' *Financial Analysts Journal*, 63(6).
- Fama, E. F. and French, K. R. (2008) 'Dissecting Anomalies.' *Journal of Finance*, 63(4) pp. 1653–1678.
- Fama, E. F. and French, K. R. (2015) 'A five-factor asset pricing model.' *Journal of Financial Economics*, 116(1) pp. 1–22.
- Fama, E. F. and MacBeth, J. D. (1973) 'Risk, Return, and Equilibrium: Empirical Tests.' *Journal of Political Economy*, 81(3) pp. 607–636.
- Fan, S., Opsal, S. and Yu, L. (2015) 'Equity Anomalies and Idiosyncratic Risk Around the

- World.' *Multinational Finance Journal*, 19(1) pp. 33–75.
- Fang, H. and Lai, T.-Y. (1997) 'Go-Kurtosis and Capital Asset Pricing.' *Financial Review*, 32(2) pp. 293–307.
- Fink, J., Fink, K. E., Grullon, G. and Weston, J. P. (2010) 'What Drove the Increase in Idiosyncratic Volatility during the Internet Boom?' *Journal of Financial and Quantitative Analysis*, 45(5) pp. 1253–1278.
- Friedman, B. M. and Liabson, D. I. (1989) 'Economic Implications Of Extraordinary Movements In Stock Prices.' *Brookings Papers on Economic Activity*, 2 pp. 137–189.
- Fu, F. (2009) 'Idiosyncratic risk and the cross-section of expected stock returns.' *Journal of Financial Economics*, 91 pp. 24–37.
- Fu, F. and Schutte, M. G. (2010) 'Investor Diversification and the Pricing of Idiosyncratic Risk.' In *Financial Management Association Asian Conference*. Research Collection Lee Kong Chian School Of Business.
- Gnedenko, B. (1943) 'Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire.' *Annals of Mathematics, Second Series*, 44(3) pp. 423–453.
- Goetzmann, W. N. and Kumar, A. (2001) *Equity Portfolio Diversification*. NBER Working Papers.
- Gordon, M. J. and Shapiro, E. (1956) 'Capital Equipment Analysis: The Required Rate of Profit.' *Management Science*, 3(1) pp. 102–110.
- Gospodinov, N. and Robotti, C. (2013) 'Asset Pricing Theories, Models, and Tests.' In Baker, H. K. and Filbeck, G. (eds) *Portfolio Theory and Management*. Oxford University Press.
- Goyal, A. and Santa-Clara, P. (2003) 'Idiosyncratic Risk Matters!' *Journal of Finance*, 58(3) pp. 975–1007.
- Granger, C. W. J. (1969) 'Investigating Causal Relations by Econometric Models and Cross-spectral Methods.' *Econometrica*, 37(3) pp. 424–438.
- Granger, C. W. J. and Newbold, P. (1986) *Forecasting Economic Time Series*. 2nd ed., Academic Press.
- Gray, D. E. (2014) *Doing Research in the Real World*. 3rd ed., SAGE Publications Ltd.
- Gray, S., Hall, J., Diamond, N. and Brooks, R. (2013) 'Comparison of OLS and LAD regression techniques for estimating beta.' *Report submitted in response to AER Rate of Return Guideline Consultation Paper*, June.
- Gunthorpe, D. and Levy, H. (1994) 'Portfolio Composition and Investment Horizon.' *Financial Analysts Journal*, 50(1) pp. 51–56.
- Guo, H., Kassa, H. and Ferguson, M. F. (2014) 'On the Relation between EGARCH Idiosyncratic Volatility and Expected Stock Returns.' *Journal of Financial and*

Quantitative Analysis, 49(1) pp. 271–296.

- Haas, M. and Pigorsch, C. (2009) ‘Financial Economics, Fat-tailed Distributions.’ *Encyclopedia of Complexity and Systems Science*. Springer New York.
- Hamilton, J. D. and Susmel, R. (1994) ‘Autoregressive conditional heteroskedasticity and changes in regime.’ *Journal of Econometrics*, 64 pp. 307–333.
- Hansen, P. R. and Lunde, A. (2005) ‘A forecast comparison of volatility models: does anything beat a GARCH(1,1)?’ *Journal of Applied Econometrics*, 20(7) pp. 873–889.
- Harvey, C. R. and Siddique, A. (2000) ‘Conditional Skewness in Asset Pricing Tests.’ *Journal of Finance*, LV(3) pp. 1263–1295.
- Hershey, J. C. and Schoemaker, P. J. H. (1980) ‘Risk Taking and Problem Context in the Domain of Losses: An Expected Utility Analysis.’ *Journal of Risk and Insurance*, 47(1) pp. 111–132.
- Huang, W., Liu, Q., Rhee, S. G. and Wu, F. (2012) ‘Extreme downside risk and expected stock returns.’ *Journal of Banking and Finance*, 36 pp. 1492–1502.
- Huang, W., Liu, Q., Rhee, S. G. and Zhang, L. (2012) ‘Return Reversals, Idiosyncratic Risk, and Expected Returns.’ *Review of Financial Studies*, 23 pp. 147 – 168.
- Ince, O. S. and Porter, R. B. (2006) ‘Individual Equity Return Data from Thomson Datastream: Handle with Care!’ *Journal of Financial Research*, 19(4) pp. 463–479.
- Iwasawa, S. and Uchiyama, T. (2013) ‘A Behavioral Economics Exploration into the “Volatility Anomaly.”’ *Policy Research Institute, Ministry of Finance, Japan, Public Policy Review*, 9(3) pp. 457–490.
- J.P.Morgan/Reuters (1996) *RiskMetrics - Technical Document*. 4th ed., New York: J.P.Morgan/Reuters.
- Jegadeesh, N. and Titman, S. (1993) ‘Returns to buying winners and selling losers: implications for stock market efficiency.’ *Journal of Finance*, 48(1) pp. 65–91.
- Jondeau, E. and Rockinger, M. (2000) ‘Conditional Volatility, Skewness and Kurtosis: Existence and Persistence.’ *HEC Paris, Cahiers de recherche No. 710*.
- Jones, C. (2001) ‘Extracting factors from heteroscedastic asset returns.’ *Journal of Financial Economics*, 62 pp. 293–325.
- Kahneman, D. and Tversky, A. (1979) ‘Prospect Theory: An Analysis of Decision under Risk.’ *Econometrica*, 47(2) pp. 263–292.
- Kan, R. and Zhang, C. (1999a) ‘GMM Tests of Stochastic Discount Factor Models with Useless Factors.’ *Journal of Financial Economics*, 54(1) pp. 103–127.
- Kan, R. and Zhang, C. (1999b) ‘Two-Pass Tests of Asset Pricing Models with Useless Factors.’ *Journal of Finance*, 54(1) pp. 204–235.

- Kandel, S. and Stambaugh, R. F. (1989) 'A Mean-Variance Framework for Tests of Asset Pricing Models.' *Review of Financial Studies*, 2 pp. 125–156.
- Keim, D. B. and Madhavan, A. (1998) 'The Cost of Institutional Equity Trades.' *Financial Analysts Journal*, 54(4) pp. 50–69.
- Kelly, M. (1995) 'All their eggs in one basket: Portfolio diversification of US households.' *Journal of Economic Behavior and Organization*, 27 pp. 87–96.
- Kenneth R. French Data Library (2015). [Online] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- Khovansky, S. and Zhylyevskyy, O. (2013) 'Impact of idiosyncratic volatility on stock returns: A cross-sectional study.' *Journal of Banking & Finance*, 37 pp. 3064–3075.
- Kingsley, P. (2012) 'Financial crisis: timeline.' *The Guardian*, August.
- Koenker, R. and Gilbert Bassett, J. (1978) 'Quantile Regression.' *Econometrica*, 46(1) pp. 33–50.
- Koenker, R. and Machado, J. A. F. (1999) 'Goodness of Fit and Related Inference Processes for Quantile Regression.' *Journal of the American Statistical Association*, 94(448) pp. 1296–1310.
- Kraus, A., Litzenberger, R. H. and Litzenberg, R. H. (1976) 'Skewness Preference and the Valuation of Risk Assets.' *Journal of Finance*, 31(4) pp. 1085–1100.
- Lee, G. J. and Engle, R. F. (1999) 'A permanent and transitory component model of stock return volatility.' In Engle, R. F. and White, H. (eds) *Cointegration Causality and Forecasting A Festschrift in Honor of Clive W.J. Granger*. Oxford University Press, pp. 475–497.
- Levin, A. E. (1995) 'Stock Selection via Nonlinear Multi-Factor Models.' In Touretzky, D. S. and Hasselmo, M. E. (eds) *Advances in Neural Information Processing Systems*.
- Levy, H. (1978) 'Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio.' *American Economic Review*, 68(4) pp. 643–658.
- Levy, H. (2012) *The Capital Asset Pricing Model in the 21st Century: Analytical, Empirical, and Behavioral Perspectives*. Cambridge University Press.
- Li, X., Sullivan, R. N. and Garcia-Feijóo, L. (2014) 'The Limits to Arbitrage and the Low-Volatility Anomaly.' *Financial Analysts Journal*, 70(1) pp. 52–63.
- Lindner, A. M. and Meyer, K. M. M. (2003) 'Extremal Behavior of finite EGARCH processes.' *Discussion Paper / Sonderforschungsbereich 386 der Ludwig-Maximilians-Universitaet Muenchen No.347*.
- Lintner, J. (1965a) 'Security Prices and Risk: The Theory of Comparative Analysis of AT&T and Leading Industrials.' *Paper presented at the Conference on Economics of Public Utilities, Chicago, 1965*.

- Lintner, J. (1965b) 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.' *Review of Economics and Statistics*, 47(1) pp. 13–37.
- Liu, H. and Loewenstein, M. (2002) 'Optimal Portfolio Selection with Transaction Costs and Finite Horizons.' *Review of Financial Studies*, 15(3) pp. 805–835.
- Lo, A. W., Petrov, C. and Wierzbicki, M. (2003) 'It's 11pm - Do You Know Where Your Liquidity Is? The Mean-Variance Liquidity Frontier.' *Journal Of Investment Management*, 1(1) pp. 55–93.
- De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990) 'Noise Trader Risk in Financial Markets.' *Journal of Political Economy*, 98(4) pp. 703–738.
- Lucas, R. E. (1978) 'Asset Prices in an Exchange Economy.' *Econometrica*, 46(6) pp. 1429–1445.
- Lyandres, E., Sun, L. and Zhang, L. (2008) 'The New Issues Puzzle: Testing the Investment-Based Explanation.' *Review of Financial Studies*, 21(6) pp. 2825–2855.
- Malkiel, B. G. and Xu, Y. (2004) 'Idiosyncratic Risk and Security Returns.' *In AFA 2001 New Orleans Meetings*.
- Markowitz, H. (1952) 'Portfolio Selection.' *Journal of Finance*, 7(1) pp. 77–91.
- Markowitz, H. M. (ed.) (1959) *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, Inc.
- Mas-Colell, A., Whinston, M. D. and Green, J. R. (1995) *Microeconomic Theory*. Oxford University Press.
- McLean, R. D. and Pontiff, J. (2016) 'Does Academic Research Destroy Stock Return Predictability?' *Journal of Finance*, 71(1) pp. 5–32.
- Meese, R. A. and Rogoff, K. (1983) 'Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?' *Journal of International Economics*, 14 pp. 3–24.
- Meese, R. A. and Rogoff, K. (1988) 'Was it Real? The Exchange Rate Differential Relation Over the Modern Floating-Rate Period.' *Journal of Finance*, 43(3) pp. 933–948.
- Merton, R. C. (1974) 'On the pricing of corporate debt: the risk structure of interest rates.' *Journal of Finance*, 29(2) pp. 449–470.
- Merton, R. C. (1980) 'On Estimating the Expected Return on the Market.' *Journal of Financial Economics*, 8 pp. 323–361.
- Merton, R. C. (1987) 'A Simple Model of Capital Market Equilibrium with Incomplete Information.' *Journal of Finance*, 42(3) pp. 483–510.
- Mikosch, T. and Starica, C. (2000) 'Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process.' *Annals of Statistics*, 28(5) pp. 1427–1451.

- Miller, M. and Scholes, M. (1972) 'Rates of Return in Relation to Risk: A Reexamination of Some Recent Studies.' In Jensen, M. (ed.) *Studies in the Theory of Capital Markets*. New York: Praeger.
- Mincer, J. and Zarnowitz, V. (1969) 'The Evaluation of Economic Forecasts.' In Mincer, J. A. (ed.) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. Washington: NBER, pp. 3–46.
- Mishkin, F. S. and White, E. N. (2002) 'U.S. Stock Market Crashes and Their Aftermath: Implications for Monetary Policy.' *National Bureau for Economic Research Working Paper*, (8992) June.
- Nasdaq (n.d.) *Financial Glossary*.
- Nelson, D. B. (1990) 'Stationarity and Persistence in the GARCH(1,1) Model.' *Econometric Theory*, 6(3) pp. 318–334.
- Nelson, D. B. (1991) 'Conditional Heteroscedasticity in Asset Returns: A New Approach.' *Econometrica*, 59 pp. 347–370.
- von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*. Princeton University Press.
- Newey, W. K. and West, K. D. (1987) 'A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix.' *Econometrica*, 55(3) pp. 703–708.
- Nobelprize.org (1990) 'The Prize in Economics 1990 - Press Release.' Nobel Media AB 2014. [Online] [Accessed on 19th July 2016] http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1990/press.html.
- Northfield (2013) 'US Macroeconomic Equity Risk Model.'
- Owen, J. and Rabinovitch, R. (1983) 'On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice.' *Journal of Finance*, 38(3) pp. 745–752.
- Pagan, A. R. and Schwert, G. W. (1990) 'Alternative models for conditional stock volatility.' *Journal of Econometrics*, 45 pp. 267–290.
- Papaioannou, M. G., Park, J., Pihlman, J. and Hoorn, H. van der (2013) *Procyclical Behavior of Institutional Investors During the Recent Financial Crisis: Causes, Impacts, and Challenges*. (Working Paper, Monetary and Capital Markets Department).
- Pompian, M. (2016) *Understanding Behavioral Biases*. CFA Institute Online Courses. [Online] <https://www.cfainstitute.org/learning/products/onlinelearning/Pages/128064.aspx>.
- Roll, R. (1977) 'A Critique of Asset Pricing Theory's Tests: Part I: On Past and Potential Testability of the Theory.' *Journal of Financial Economics*, 4 pp. 129–176.
- Roll, R. (1984) 'A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market.' *Journal of Finance*, 39 pp. 1127–1139.

- Ross, S. A. (1976) 'The Arbitrage Theory of Capital Asset Pricing.' *Journal of Economic Theory*, 13 pp. 341–360.
- Ruan, T., Sun, Q. and Xu, Y. (2010) 'When Does Idiosyncratic Risk Really Matter?' *In The Fifth Annual Conference on Asia-Pacific Financial Markets*. Korean Securities Association.
- Ryan, B., Scapens, R. W. and Theobald, M. (2002) *Research Method and Methodology in Finance and Accounting*. 2e ed., Cengage Learning EMEA.
- Salmon, M. H., Earman, J., Glymour, C., Lennox, J. G., Machamer, P., McGuire, J. E., Norton, J. D., Salmon, W. C. and Schaffner, K. F. (1999) *Introduction to the Philosophy of Science*. Indianapolis: Hackett Publishing Company.
- Saunders, M., Lewis, P. and Thornhill, A. (2009) *Research methods for business students*. 5e ed., Pearson Education.
- Scowcroft, A. and Sefton, J. (2005) 'Understanding Momentum.' *Financial Analysts Journal*, 61(2).
- SEC (n.d.) *Division of Trading and Markets*.
- Sharpe, W. F. (1964) 'Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.' *Journal of Finance*, 19(3) pp. 425–442.
- Shiryaev, A., Xu, Z. and Zhou, X. Y. (2008) 'Thou shalt buy and hold.' *Quantitative Finance*, 8(8) pp. 765–776.
- Sloan, R. G. (1996) 'Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?' *Accounting Review*, 71(3) pp. 289–315.
- Smith, V. L., Suchanek, G. L. and Williams, A. W. (1988) 'Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets.' *Econometrica*, 56(5) pp. 1119–1151.
- Spiegel, M. I. and Wang, X. (2005) 'Cross-sectional Variation in Stock Returns: Liquidity and Idiosyncratic Risk.' *Yale ICF Working Paper No. 05-13; EFA 2005 Moscow Meetings Paper*, September.
- Statman, M. (1987) 'How Many Stocks Make a Diversified Portfolio.' *Journal of Financial and Quantitative Analysis*, 22(3) pp. 353–363.
- Stewart, R. B. (2010) *Value Optimization for Project and Performance Management*. Wiley.
- Subrahmanyam, A. (2007) 'Behavioural Finance: A Review and Synthesis.' *European Financial Management*, 14(1) pp. 12–29.
- Taramasco, O. and Bauer, S. (2013) "'RHmm": Hidden Markov Models simulations and estimations [software], version 2.0.3.'
- Tobin, J. (1956) 'Liquidity Preference as Behavior Towards Risk.' *Cowles Foundation*

Discussion Paper No. 14, July.

- Tobin, J. (1958) 'Liquidity Preference as Behavior Towards Risk.' *Review of Economic Studies*, 25(1) pp. 65–86.
- Violante, F. and Laurent, S. (2012) 'Volatility Forecasts Evaluation and Comparison.' In Bauwens, L., Hafner, C., and Laurent, S. (eds) *Handbook of Volatility Models and Their Applications*.
- Viterbi, A. J. (1967) 'Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm.' *IEEE Transactions on Information Theory*, VOL. IT-13, NO. 2, APRIL 1967, IT-13(2) pp. 260–269.
- Walkshäusl, C. (2013) 'The high returns to low volatility stocks are actually a premium on high quality firms.' *Review of Financial Economics*, 22(4) pp. 180–186.
- Woodward, J. (2014) *Scientific Explanation*. The Stanford Encyclopedia of Philosophy.
- Yin, R. K. (2011) *Qualitative Research From Start to Finish*. The Guilford Press.