Financial Risk Management

From a Global to an Individual Perspective

Rik Frehen
Financial Risk Management
From a Global to an Individual Perspective

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 22 januari 2010 om 12.00 uur

door

Rik Guus Phil Frehen
Promotores

Prof. dr. R.M.M.J. Bauer
Prof. dr. P.C. Schotman

Beoordelingscommissie

Prof. dr. P.M.A. Eichholtz (voorzitter)
Dr. W.F.M. Bams
Prof. W.N. Goetzmann (Yale University)
Prof. dr. F.C.J.M. de Jong (Tilburg University)
Prof. dr. F.C. Palm

I thank Inquire Europe, The Maastricht Research School of Economics of Technology and Organizations (METEOR), The Network for Studies on Pensions, Aging and Retirement (NETSPAR), and Yale University for financial support.
Acknowledgements

Almost five years ago I entered Rob’s office, asking him if there would be an opportunity for me to pursue a PhD. Rob was enthusiastic and promised me a position, while Peter worked hard to finance it. I am still very grateful for these efforts. Due to logistical constraints, it was not until July 2005 that I started, while teaching at a high school in the mean time.

From the outset I have enjoyed doing research and teaching. Nevertheless, I have experienced volatile times and would never have been able to finish my thesis without the help of many others. Therefore I would like to express my gratitude towards them.

First, I would like to thank my supervisors. Rob, you supported me during the first stage, when I badly needed it and you have shown confidence in me. Peter, I have learned from you, especially during seminars and paper discussions. Your constructive criticism has been very valuable. The both of you have also given me the freedom to develop myself the way I did. I would also like to thank Piet for his advice in some of my career decisions and the members of the reading and assessment committee, for carefully reading my thesis.

I have also benefited from our finance department, Inquire Europe, Netspar and Yale University through financial support. I am very grateful for financing my trips and research. I realize that I have been extremely lucky with the strong financial backing, which enabled me to visit many conferences and go abroad for a while. These opportunities have been crucial for my development during the last four years.

Special thanks to my co-authors at Yale: Geert, Martijn and Will. I had a wonderful time in New Haven and still learn so much from you. It is a true pleasure working with you and I am very much indebted to you for all the opportunities you provided.

In addition, I would like to thank my friends at the university: Frank, Jeroen, Kasper, Robin and Stephan for the great time I had with you. I really appreciate all the lunches and non-research-related discussions. You definitely made it a lot easier and more fun for me to finish this dissertation. Special thanks also to my paranymphs, Dion and Mathijs. Dion, I thank you for spending so many great times together (during our studies and in New Haven) and for your always helpful advice. Mathijs, I maybe learned most from you and am very grateful for your friendship. It was fantastic to share a room with you and we had great fun in
ACKNOWLEDGEMENTS

Maastricht and during our trips to conferences.

De grootste dank gaat uit naar mijn familie. Loes en Barry, hartelijk bedankt voor jullie constante interesse en vertrouwen. Pap en mam, ik wil jullie bedanken voor jullie onvoorwaardelijke steun, liefde en begrip, die veel verder reiken dan de laatste vier jaar. Jullie hebben me gemaakt tot wie ik ben en zonder jullie had ik dit nooit bereikt. Katrin, je raad, je begrip, je luisterende oor en je geloof in mij, zijn nog maar een paar van alle dingen waar ik je voor wil bedanken. Zonder jou waren de laatste jaren niet half zo mooi geweest. Ik ben trots op wat wij samen hebben.

Rik Frehen
Sittard, August 2009
Contents

1 Introduction 1
1.1 Global financial risk and bubbles 2
1.2 Measuring risk of individual stocks 3
1.3 Risk-adjusted Equity Performance of Pension Funds 4
1.4 Pricing Retirement Risk Protection and Behavioral Biases 5

2 New Evidence on the First Financial Bubble 7
2.1 Introduction 7
2.2 Background 9
2.3 Data 11
2.4 Cross-sectional Differences in the South Sea Bubble 12
2.5 International Stock Market Crash 17
2.6 Cross-sectional Evidence from New Company Issues 20
2.7 The IPO Wave in the Netherlands 22
2.8 Conclusion 25
2.A Appendix I: The British Bubble 26
2.B Appendix II: The Dutch Bubble 29
2.C Appendix III: Price Quote Interpretation 30

3 Efficient Estimation of Firm-Specific Betas and its Benefits for Asset Pricing Tests and Portfolio Choice 33
3.1 Introduction 33
3.2 The Model 37
3.3 Methodology 40
3.3.1 Bayesian Methods 40
3.3.2 Prior Distributions 40
3.3.3 Bayesian Inference 41
3.4 Data 42
3.5 Empirical Results 44
3.5.1 Beta Specification 44
3.5.2 Beta Estimation 48
3.5.3 Portfolio Heterogeneity 53
3.6 Applications of Firm-Specific Betas 56
Chapter 1

Introduction

The recent crisis has shown that financial risk management not only lies at the core of our economy, but at the heart of our modern society. Increasingly strong links between international markets and various sectors in our economy make the quantification and monitoring of financial risks essential for the health of our financial system and thus for the functioning of the entire society. Financial risk affects all the things which are most precious to all of us: our health care system, our retirement date and income, the overall quality of education and even the performance of our favorite football club. As a consequence, measuring, monitoring and managing risk have become increasingly important and the field of finance is the axis in many decision making processes.

The previous examples illustrate the importance of financial risk management, but they also show how broad the field is. Addressing all types of financial risk is not only beyond the scope of this thesis, but probably even a mission impossible. We therefore only address certain types of risk and try to narrow down the focus of this thesis per chapter. We start on a macroeconomic level, by determining the components of global financial risk and more specifically international bubbles in equity prices. We proceed by focusing on the risk of individual stocks in the US. We measure the risk in terms of CAPM betas and show the effects of beta estimates on asset pricing tests and portfolio choice. Thereafter, we narrow our universe further down by looking at domestic equity investment by US pension funds. The risk-adjusted performance of domestic equity portfolios is measured and we give insights in the benchmarks for performance evaluation and cost levels of the pension industry. Eventually, we focus on the individual pension plan participant, by measuring the annuity risk that he faces at the retirement date. We develop and price an option that protects participants against this risk, taking into account possible regret in the decision making process.
1 Introduction

1.1 Global financial risk and bubbles

Bubbles in equity prices have attracted enormous scholarly interest ever since the infamous Dutch tulip mania in 1637. Over the last centuries many efforts were initiated in trying to understand the dynamics of the bubbles. The literature on the topic may be best described by the most extreme views on these extraordinary events; irrational exuberance\(^1\) on the one hand and fully rational expectations\(^2\) on the other. Despite the disagreement on the rationality of the bubble, there seems to be consensus on the importance of innovation in many bubbles. Examples of the link between bubbles and innovation are the IT innovations in the nineties and the corresponding NASDAQ bubble and the emergence of railroad transportation with the accompanying railroad mania halfway the nineteenth century. Based on this last example a framework was developed for explaining the bubbles in share prices during innovations.\(^3\)

The first global financial bubble in stock prices occurred in 1720 in Paris, London and the Netherlands and was never considered to be caused by innovations. Explanations for these linked bubbles primarily focus on the irrationality of investor speculation and the corresponding stock price behavior of two large firms: the South Sea Company in Great Britain and the Mississippi Company in France. In chapter 2 we examine a broad cross-section of largely new security price data to evaluate the causes of the bubbles. Using the collected stock prices for British and Dutch firms in 1720, we find evidence against indiscriminate irrational exuberance and evidence in favor of speculation about two factors: the Atlantic trade and the incorporation of insurance companies.

The war of the Spanish Succession and the war of the Quadruple Alliance had severely weakened the Spanish position in the Americas and deprived them of the Asiento, the right to trade slaves with the West. Expectations about future profits of the Atlantic trade led to price run ups in the firms trading with the West. Beside the developments overseas, a financial innovation in the insurance business took place. From pooling private capital for the insurance of ships by rich business men, insurance gradually transformed into a new corporate form. Limited liability companies were set up, gathering large pools of capital via the issuance of shares. Through many discussions of politicians and business men, we are able to determine that the expectations about the future profits of the new corporate form were high. Consequently, the share prices of insurance companies bubbled much stronger than those of any other industry. Our findings are therefore consistent with the hypothesis that financial bubbles require a plausible story to justify investor optimism.

\(^1\) Cf. Dale, Johnson, and Tang (2005) and Vissering (1863)
\(^2\) Cf. Garber (1990)
\(^3\) Cf. Pastor and Veronesi (2009)
1.2 Measuring risk of individual stocks

Ever since the Capital Asset Pricing Model was introduced by among others Sharpe (1964) and Lintner (1965), it has been a cornerstone in the asset pricing literature. During the last decades many extensions and improvements on the CAPM were proposed. One of the most important generalizations concerns the time variation in betas. Although many empirical studies provide evidence of the time variation in betas, there is no agreement on the best specification. The literature is sharply divided in parametric and non-parametric specifications for beta dynamics. In the parametric specification\(^4\) betas are a function of conditioning variables, whereas non-parametric specifications use rolling windows regressions, short window regressions or Kalman-filters to model the time variation. Another debate in the asset pricing literature concerns the estimation methods. A majority of studies groups stocks into portfolios when estimating asset pricing models\(^5\), whereas others\(^6\) argue that individual stocks should be used for asset pricing purposes.

In chapter 3, we improve both the specification and estimation of firm-specific betas. Time variation in betas is modeled by combining a parametric specification based on economic theory with a non-parametric approach based on data-driven filters. The non-parametric specification is improved by mixing data at different frequencies using the MIDAS-approach, and by specifying a flexible weighting function that gives larger weights to more recent observations. We allow the optimal combination of parametric and non-parametric betas to vary per stock and over time. We also increase the precision of individual beta estimates by setting up a hierarchical Bayesian panel data model that imposes a common structure on the parameters. The model shrinks the firm-specific beta estimates towards the cross-sectional mean for stocks with few time series observations. The hierarchical structure also enables us to include a large number of factors in our model, while efficiently estimating the parameters of the hierarchical distribution.

We show that the dynamics of large, value stocks are better described by a non-parametric specification, whereas for small growth stocks the parametric approach is preferred. Furthermore, the optimal specification loads more heavily on the stable, non-parametric beta during periods of distress. We also document that our accurate beta estimates lead to a large increase in the cross-sectional explanatory power of the conditional CAPM. Our new specification results in sharply reduced and insignificant pricing errors, loads significantly on the beta estimates from the panel model in cross-sectional regressions and gives risk premium estimates very close to the sample mean. We are also able to confirm the finding by Ang, Liu, and Schwarz (2008), that the formation of traditional asset pricing portfolios leads to a reduction in cross-sectional variation in betas, which results in large standard errors of the constant and relatively small loadings on betas in cross-sectional regressions. In addition, we provide evidence that there is considerable variation in alphas and betas within traditional asset pricing portfolios. Finally, chapter 3

\(^{4}\) Cf. Shanken (1990) and Ferson and Harvey (1999)

\(^{5}\) Cf. Fama and MacBeth (1973) and Fama and French (1992)

\(^{6}\) Cf. Avramov and Chordia (2006a) and Ang, Liu, and Schwarz (2008)
shows that using the betas to forecast the covariance matrix of returns results in a significant improvement in the out-of-sample performance of minimum variance portfolios.

1.3 Risk-adjusted Equity Performance of Pension Funds

The risk measurement and management of pension funds is one of the most challenging tasks in current day finance. Asset allocation, asset-liability management, longevity risk and intergenerational risk sharing are just a taste of the many decisions and risks at stake for pension fund managers. Strangely enough, when compared to the mutual fund industry, pension fund managers enjoy about half of the support from the academic world.\footnote{A search on scholar.google.com for “mutual fund” results in 708,000 hits, whereas “pension fund” leads to only 359,000 hits. Similarly, a search on sciencedirect.com for “mutual fund” results in 24,665 articles and “pension fund” in 12,041 articles.} The low number of academic studies on pension funds can be attributed to a general lack of comprehensive data on pension funds. Where mutual funds have an obligation to report their returns, investment styles and fees to the SEC on a regular basis, pension funds have no such requirement. As a result, we have a large database with mutual fund data (CRSP) and hardly any data on pension funds. It is therefore not surprising that there is general consensus on mutual fund underperformance\footnote{Cf. Malkiel (1995) and Elton, Gruber, and Busse (2004)}, while there is a large debate on the performance of pension funds\footnote{Cf. Busse, Goyal, and Wahal (2009) and Lakonishok, Shleifer, and Vishny (1992)}.

The different conclusions on pension fund performance may simply be attributed to the various characteristics of the data sets employed in earlier work. Some studies focus their attention on defined benefit funds (which are generally larger), whereas others merely look at defined contribution plans. Variation in performance figures may also be explained by differences between corporate, public or other\footnote{e.g. churches and universities} funds. Data sets typically contain a specific type of fund, e.g. corporate defined benefit funds, which make generalizations of results disputable. A more fundamental problem of pension fund data concerns the data provider. A large majority of studies is performed on manager-level data, i.e. data provided by asset management companies, investing on behalf of pension funds. Although this may give insight in the overall performance, scholars are not able to distinguish between funds since one manager may work for more than one fund, and funds employ more than one manager. Finally, all pension data provided by either funds or asset management companies may suffer from biases since there is no obligation to report.

CEM Benchmarking Inc. (CEM) was one of the first to recognize the difficulties with pension fund data and started to send out questionnaires to funds from 1990 on, requesting performance, benchmark and cost figures. CEM offered pen...
1.4 Pricing Retirement Risk Protection and Behavioral Biases

Narrowing our center of attention further down to the individual, we focus on the defined contribution pension plan participant shortly before his retirement date. In the years before retirement, defined contribution plan participants face considerable risks. Investment risk can be reduced by gradually moving towards a riskless strategy as the retirement date approaches. Mortality risk can be covered by standard insurance products. However, annuity risk is virtually impossible to hedge and yet can affect the after-retirement income drastically. As a result, defined contribution plan participants confront a difficult decision shortly before their retirement date: if and when should they convert their retirement capital

\[11\] This last distinction is only possible for defined benefit funds due to data restrictions.

\[12\] Cf. Swensen (2005)
into a life annuity? The high duration of the life annuity induces a substantial price sensitivity with respect to long-term interest rates. Retiring a month earlier or later can therefore easily amount to 10% difference in after-retirement income. Consequently, the option to time the retirement date is extremely valuable.\textsuperscript{13} In addition to the risks and uncertainties of the retirement investment decision, numerous experiments have shown\textsuperscript{14} that regret aversion plays an important role in human decision making, complicating the process even further. Imagine that your co-worker, who retired half a year earlier than you, has a 10% higher income, while converting the same amount of capital at retirement.

In order to protect defined contribution plan participants against annuity risk at their retirement date and possible regret of bad timing, we develop and price a lookback option on a life annuity contract in chapter 5. The lookback life annuity option guarantees that participants have made the optimal timing decision and provides the ultimate protection against annuity risk at retirement. We first determine a closed-form option value under the restriction that the option holder invests risklessly during the time to maturity of the option and without the guarantee that the exact amount of retirement capital is converted into a life annuity at retirement. Thereafter, the investment restriction is relaxed, the guarantee of exact conversion is imposed and the option is priced via Monte Carlo simulations in an economic environment with a stochastic discount factor. We proceed by determining the sensitivities of the option price with respect to its characteristics, i.e. the type of option, initial interest rate, time to maturity and asset allocation.

We find that the price of a lookback option, with a maturity of three years, amounts to 8-9% of the capital at the option issuance date. The option price is most sensitive to the exercise price of the option, i.e. pricing alternative options (e.g. Asian) substantially lowers the price. Time to maturity and interest rate volatility are other important option price drivers. Asset allocation decisions and initial interest rates hardly affect the option price.

Finally, some further remarks on the structure of the thesis. In some cases, we refer for more information to the appendix of the chapter. Further explanations are always provided in the appendix of the chapter for which the information is relevant. Typically, we refer to appendices for proofs, posterior distributions, derivations and additional material. The notation used throughout this thesis, although fairly similar, may differ for each chapter, but within each chapter the notation is consistent. Relevant notation is therefore defined within the chapters.

\begin{flushright}
\textsuperscript{13} Cf. Milevsky and Young (2006)
\textsuperscript{14} Cf. Loomes and Sugden (1982) and Kahneman and Tversky (1979)
\end{flushright}
Chapter 2

New Evidence on the First Financial Bubble

The first global financial bubble in stock prices occurred in 1720 in Paris, London and the Netherlands. Explanations for these linked bubbles primarily focus on the irrationality of investor speculation and the corresponding stock price behavior of two large firms: the South Sea Company in Great Britain and the Mississippi Company in France. In this chapter we examine a broad cross-section of security price data to evaluate the causes of the bubbles. Using newly collected stock prices for British and Dutch firms in 1720, we find evidence against indiscriminate irrational exuberance and evidence in favor of speculation about two factors: the Atlantic trade and the incorporation of insurance companies. The fundamentals of both sectors may have led to high expectations of future growth. Our findings are consistent with the hypothesis that financial bubbles require a plausible story to justify investor optimism.¹

2.1 Introduction

Asset bubbles are important puzzles in financial economics - important because of their extraordinarily potential for disruption; puzzles because they defy standard notions of rationality. Recent research in the wake of the NASDAQ bubble and the global housing bubble has highlighted the role of uncertainty and technological innovation in the development of asset bubbles. In particular, Pastor and Veronesi (2009) and Pastor and Veronesi (2006) argue that technological innovation and investor uncertainty can explain high security valuations and the rapid rise and fall of stocks related to potentially transformative technological innovations. Their papers make cross-sectional empirical predictions about securities during periods of technological change. They show that imputed growth rates in

¹ This chapter is based on the paper Frehen, Goetzmann, and Rouwenhorst (2009).
innovative industries can appear irrationally high ex post and that industries in which bubbles occur will be characterized by high return volatility, high uncertainty and rapid adoption of the new technology. They test these predictions on 19th century railroad securities listed on the New York Stock Exchange.

In this chapter we revisit one of the most famous events in financial history: the South Sea Bubble of 1720. Along with the earlier Dutch Tulipmania, of the 17th century, the South Sea Bubble is often regarded as an example par excellence of investor irrationality. However, unlike Tulipmania, which was confined to the futures market in tulip bulbs, the South Sea Bubble involved many stocks traded in three different countries. Although 1720 is not generally viewed as a period of technological novelty, we argue in this chapter that there were at least three critical innovations that took place in a very short span of time: two of which were financial innovations, the other was a major potential shift in the configuration of global trade. The first innovation was financial engineering at a national scale. The Mississippi Company and the South Sea Company issued equity shares in exchange for government debt; in effect converting the national debt into corporate stock. This novelty appeared to be a new model for government finance: a heavily politically-influenced corporation that also had exclusive rights and patents to pursue other ventures. The second innovation was an incipient shift in global trade. Both of the companies were set up to exploit trade in the Americas. The Mississippi Company owned the Louisiana territory and the South Sea Company owned the exclusive right to export African slaves to Spanish America. Spanish domination of the Atlantic trade was weakened as a result of the War of the Spanish Succession [1701-1714], and the War of the Quadruple Alliance [1718-1720], opening the door to increased competition from Britain, France and the Netherlands. These geopolitical conditions offered great possibilities, but also great uncertainty.

The third innovation was also financial. The first publicly traded insurance corporations were chartered in Great Britain 1720, as a result of the famous Bubble Act. As such, they represented a new model of capital formation and risk-sharing for maritime insurance firms - in a nation built on maritime trade. This innovation was rapidly imitated in the Netherlands, despite past Dutch dominance of maritime insurance. The innovation in insurance form also stirred great debate and uncertainty about its effects.

Ex post, two of these three innovations had a transformative effect on the global economy. Although government debt for equity swaps did not survive the financial crisis when the South Sea bubble burst, the Atlantic trade did. The triangle trade between Europe, Africa and the Caribbean eventually became the dominant international trade system of the 18th and early 19th centuries. Finally, the emergence of publicly financed, limited liability insurance companies ultimately transformed risk sharing. The traditional insurance industry, consisting of underwriters which pooled capital collected via brokers, encountered fierce competition from the new corporations. The publicly traded insurance companies founded in 1720 survived and flourished and became models for the insurance trade in both the old and the new worlds. The public insurance corporation is now the dominant institutional
form. In this chapter we examine the extent to which the bubble of 1720 might be explained by current models of technological innovation and share price dynamics. To do so, we collect cross-sectional security price information for a number of stocks traded in London and the Netherlands. Some of the security price data is brand new to scholarship. We use these new data to measure cross-sectional differences in the timing, steepness and magnitude of share price increases in 1720.

We find strong evidence that the dynamics of the South Sea Bubble differed by industry. We next document the extent to which these industry effects extended across national boundaries. Our results suggest that speculation about the Atlantic trade with the Americas was an important factor in investor expectations. We also find evidence that market prices and new issues in Britain and the Netherlands were driven in part by investor expectations about the financial innovations in the insurance trade. On the other hand, we find little evidence that the debt-conversion function of the Mississippi and South Sea companies was an important factor. Finally, we find some qualified support for the hypothesis that investors were simply enthusiastic about stocks in general. Most shares rose in the British and Dutch bubbles, but the growth in those not associated with the Atlantic trade or with the insurance trade was significantly less.

2.2 Background

Most early writers treat the bubbles of 1720 as evidence of investor irrationality.\(^2\) Dale, Johnson, and Tang (2005) explicitly test this theory using violations of the law of one price.\(^3\) Others have argued that the bubbles were not necessarily evidence of irrational expectations on the part of investors, but rather the result of plausible expectations of future returns\(^4\), or a reasonable response to such financial innovations as the liquidity of shares and the potential of companies to exploit new capital markets and investment opportunities.\(^5\) Garber (1990) shows how the high price for the South Sea Company might have been justified by plausible expected returns on invested capital. On the other hand, Velde (2009) comes to opposite conclusions for the Mississippi Company, arguing that in the short-term it was financially over-extended.

An important challenge for financial historians is to understand how investor expectations at the time could have driven stock prices up by many multiples in the space of a few months. What could have made investors suddenly willing to pay so much for shares in these companies? In this chapter we do not seek to test the rationality of investors but rather to provide more information about what factors may have influenced their beliefs. For example, if the bubbles in the South Sea Company and the Mississippi Company were driven by expectations about the future profitability of Atlantic trade - in which both firms held patents - then

---

\(^{2}\) Cf. Mackay (1852), Vissering (1863), Van Rijn (1899) and Kindleberger (1978).


\(^{4}\) Garber (1990).

\(^{5}\) Cf. Scott (1910) and Neal (1990).
other firms engaged in Atlantic trade should also have experienced high price run-ups in 1720, and conversely, prices of firms not engaged in the trade should not have grown as much. If the speculation was about the debt-conversion operations of the companies then other financial firms should have experienced comparable price run-ups. Thus, the cross-sectional variation in stock returns of the 1720s may reveal the factor-specific basis for investor aspirations that created the bubble - whether rational or not. The contrary hypothesis is that irrational exuberance in 1720 might simply have been indiscriminate, affecting all stocks. In this case we would not expect to find systematic differences across industries.

The barrier to examining the 1720 bubble in cross-section has always been one of data. Up to this point, the prices for some very important companies such as the Dutch East Indies and Dutch West Indies companies have not been available to researchers. Scholars have also not had access to prices for the considerable number of initial public offerings from the London and Dutch markets that were floated in 1720. While excellent British price data for several companies has been collected and studied\textsuperscript{6} no one has yet made a comparative study of the prices for British insurance companies for that period, despite the fact that they were the focus of the main Parliamentary legislation regulating financial markets in 1720, commonly known as the Bubble Act.

In this chapter we collect share prices for Dutch and English companies from a Dutch newspaper published in 1720, the Leydse Courant. We also collect price data for Royal Exchange Assurance and the London Assurance from the leading London financial periodical of the time, The Course of the Exchange. Together with the British data from Neal (1990) we are able to track the course of 26 Dutch company stocks through the entirety 1720, including the Dutch East and West Indies Companies and a broad sample of British firms over the same interval. We also use Dutch company archives to provide institutional details about mechanisms of share speculation.

We find evidence against indiscriminate irrational exuberance and evidence in favor of speculation about two factors: the Atlantic trade and the incorporation of insurance companies. The fundamentals of both sectors may have led to high expectations of future growth. The Atlantic trade was the focus of considerable political and economic activity around 1720 and the insurance business was undergoing a transition towards the corporate form. Other sectors, such as the financial industry and firms engaged in the Asian trade evidently did not inspire the same scale or investor enthusiasm. The South Sea and Mississippi bubbles were clearly not limited to major firms operating to refund government debt.

The main contribution of this chapter is the introduction of a cross-sectional analysis of share prices in 1720 in order to test competing theories about investor expectations in the first great crash of the world’s stock markets. It also introduces a hitherto unstudied source of Dutch company prices that documents the scale and timing of the crash in the Netherlands. This data set allows us to study an early and important example of the international propagation of a financial crisis.

2.3 Data

We collected prices reported for Dutch and British companies from the Leydse Courant (preserved in the National Library of the Netherlands in the Hague) over the period November 1719 to December 1720. These include London transactions of British companies, as well as occasional prices for the same firms on exchanges in the Netherlands. We added these prices to those collected by Larry Neal and used in Neal (1990) to study capital market integration in the 18th century. We augmented the Neal data with daily quotation series from the The Course of the Exchange for the two major British insurance companies.

Evidently no previous scholarship had looked at a complete sequence of the Leydse Courant, and no study so far has used regular Dutch price information for the period 1720 to 1723, although scholars had been able to gather occasional references to prices. British stock prices are available for the late 17th and early 18th century in Castaing’s The Course of the Exchange and two other sources, but Dutch data in general has been lacking for this turbulent period. The British insurance company prices in 1720 have been used to construct market indices, but not studied separately.

Leydse Courant prices were not quoted in currency but as percentages of par value net of paid-in capital. Share issues at the time were offered as subscriptions that required an initial payment that secured the subscription rights. In effect, issuing companies financed the purchase of their shares with a loan, using the shares as collateral. The initial payment was typically a small fraction (1% to 10%) of the par value of the share and came with obligations to make future payments to the company over a period of time - sometimes on a monthly basis, but other times at irregular intervals. This convention made comparisons across companies straightforward, since it had the effect of normalizing the prices. We were able to verify this reporting convention by matching share loan transactions in the books of one of the Dutch firms founded in 1720, Maatschappij van Assurantie, Discontering en Belening der Stad Rotterdam [hereafter Stad Rotterdam, its modern name], with share quotes in the Leydse Courant. Appendix 2.C discusses

---

7 Our database, including exchange rates as well as equity quotes is available online at: http://icf.som.yale.edu/DutchData/index.shtml. More data including price quotes in different cities is available upon request.
8 http://webapp.icpsr.umn.edu/cocoon/ICPSR-STUDY/01008.xml
or http://www.le.ac.uk/hi/bon/ESFDB/NEAL/neal.html
9 See Gelderblom and Jonker (2009). In a subsequent revision (2009), concurrent with the first draft of own work, they report Dutch prices for the second half of 1720. Neal (1988) mentions the Leydse Courant, but notes that only a few issues are extant for the period. In fact, he implores scholars to search Dutch libraries for a complete run of them.
10 Cf. Parsons (1974)
11 Dale, Johnson, and Tang (2005) and Shea (2007) disagree on the question of whether South Sea Company investors were obligated to continue loan payments when shares dropped in value.
this interpretation in some detail.\footnote{Although the convention is useful for interpretation of the economic scale of the events of 1720 in the Netherlands, it does not significantly affect the estimates of the bubble in share prices, since these are calculated as percentages normalized either to previous prices or previous quotes.}

By similar means we were also able to verify the assertion by Shea (2007) that, at least in Holland, subscribers in shares were not obligated to pay the firm if they chose to forego the shares. Shareholder subscription obligations were closed out when shares were returned to the company in lieu of payment.

Prices in the Leydse Courant were often quoted as a daily range. In these cases, we took the average of the range as the daily price. Prices for companies were also quoted in more than one city. In these cases we used the prices for the market with the most liquidity, defined as the market with the largest number of quotes during the period of study.\footnote{Our database preserves the range and prices for multiple markets, allowing future research analysis of such things as intra-day volatility and information diffusion and investor bias among cities, however these topics are not the focus of the current study.} Prices for the major British companies: The Bank of England, the South Sea Company, the East India Company, the Royal Exchange Assurance Company, the London Assurance Company, the York Building Society and a few others were also quoted, suggesting that Dutch investors followed - and likely traded in - British shares. This is consistent with the hypothesis in Neal (1990) that Dutch investment capital flowed into Great Britain in 1720, pursuing equity investment opportunities. News of the finances of France was also regularly reported in the paper, and prices for the shares and related financial claims on the Compagnie des Indes were quoted frequently. This suggests that Dutch investors were interested in the daily fluctuations of the French firm and may have been actively investing, at least in the year 1720. We are lacking the price information before November 1719, so it is not possible to trace earlier Dutch investor interest through this periodical.

The Leydse Courant also contains extensive interest rate and exchange rate information on the same frequency. It quoted the agio (discount rate between currency and money of account) for the Bank of Amsterdam, and exchange rates between Amsterdam and a number of other cities, including London, Paris, Lisbon and Hamburg. These are useful for future research as well.

2.4 Cross-sectional Differences in the South Sea Bubble

Figure 2.1 shows the stock price growth of the eight major London companies regularly quoted in The Course of the Exchange and the other major price list, Freke’s Price of Stocks etc. over the period from November, 1719 through December, 1720. The scale is logged to represent percentage changes and indexed to 1 at the beginning of the available quotations for each series.\footnote{Quotes for two companies, Royal Exchange Assurance and London Assurance begin January 1st, 1720 under different names.} There are three
striking features of the graph. First, during the year 1720 all firms experienced, at a minimum, a doubling of their share price. However, for three companies the prices at the end of the year were at or below their beginning of year levels. For these firms the increase in share price during the first part of 1720 was a purely temporary phenomenon, while for the others the run-up had a permanent component. Secondly, there is considerable cross-sectional variation in the degree of price run-up. Although the events of 1720 are associated with the South Sea Company, a few firms "bubbled" more dramatically than the South Sea Company, when measured by price growth. In particular, the two marine insurance companies, Royal Exchange Assurance and London Assurance rose to much higher multiples of their original price during 1720. The Royal African Company (which, like the South Sea Company, was engaged in the Atlantic slave trade) rose as high as the South Sea Company in percentage terms as well. By contrast, the two banks, Bank of England and Million Bank, as well as the East India Company (engaged in the South Asia trade) rose much less than the other four. Our limited price data for the York Building Company, a scheme that involved the development of waterworks in London, the purchase of confiscated estates in Scotland and the sale of life insurance and annuities, we cannot measure the increase over the first part of the year, although we know from other evidence that the extent of its rise from its initial offering price was similar to that of the marine insurance companies.

A final feature of interest in Figure 2.1 is the timing differences in the bubbles. The two marine insurance companies and the York Building Company reached their peak simultaneously in mid-August of 1720 and then dropped sharply at least two weeks before the crash spread to other major British firms.

The differences in scale and timing suggest multiple influences on investor enthusiasm during 1720. Clearly not all company shares benefitted equally from speculative demand. While it is difficult to draw precise conclusions from a small sample of firms, it is natural to attribute the difference between the high growth of the South Sea Company and Royal African on the one hand, and the modest growth of the East India Company on the other, to the expectations about the Atlantic trade as opposed to the South Asian trade. The new Dutch price data give us an opportunity to test this. If there were different causes of the price run-ups in Figure 2.1, reflecting differential expectations of investors about the relative fortunes of companies trading with the East versus the West, then these differences are likely to show up in the Dutch market as well. On the other hand, if the price run-ups of share prices in London were largely idiosyncratic, it is unlikely to find a similar pattern overseas.

Figure 2.2 shows the rise and fall of investment in the Dutch East Indies Company and the West Indies Company over the period of November, 1719 through December, 1720. The data are reported three times per week, so unlike the daily London prices, they are discontinuous, and represented by markers in the figure rather than by lines. The similarities between the Dutch and London markets are striking. First, the run-up in the Dutch market was much larger for the West In-

---

15 Since quotations for the York Building Society only begin in June, we do not know whether it finished up or down for the year.
2 New Evidence on the First Financial Bubble

Figure 2.1: Magnitude of the London Bubble

This figure displays the magnitude of the 1720 bubble for the Bank of England, South Sea Company, London Assurance, Royal African Company, Million Bank, York Building Company, Old East India Company and Royal Exchange Assurance. Share prices are normalized by their initial value and plotted on a logarithmic axis for the period August, 1719 to December, 1720.
2.4 Cross-sectional Differences in the South Sea Bubble

Figure 2.2: Dutch East and West Indies Company

This figure displays the share price development of the Dutch East and West Indies Company from November, 1719 to December, 1720. Prices are denoted in percentages and normalized by dividing by the first observation, plotted on a logarithmic axis and extracted from the Leydse Courant.
dies Company than the East Indies Company. Second, the relative top-to-bottom variance of prices is strikingly similar in both markets. The Dutch West Indies Company clearly bubbled on the same scale as the South Sea Company and the Royal African Company, rising by a factor of 7 over a very short interval. By comparison, the price of the Dutch East Indies Company did not double and, like its British counterpart it fell below its beginning of year value by the end of 1720.

The four major firms in Northern Europe engaged in Atlantic trade; the Mississippi Company, the Royal African Company, the South Sea Company and the Dutch West Indies Company all rose by factors of 7 to 10 in the global bubble. The two major firms engaged in Asian trade, the Dutch and British East India companies, rose by much less. This is evidence in favor of the hypothesis that investor sentiment differed with respect to Atlantic vs. Asian trading companies, supporting the theory that the South Sea Bubble may have been partly a function of expectations about future Atlantic trade. There is some supporting historical evidence for this hypothesis. The triangle trade that brought manufactured goods to Africa, Africans as slaves to the New World and plantation-produced commodities to Europe was one of the most important international economic institutions in the early modern era. It was just getting underway in 1720 and thus future profits were a long way off at the time of the bubble. Nevertheless, like the modern tech bubble, perhaps this future promise sparked current investor enthusiasm.

Other international events at the time might have directed investor attention to the Atlantic trade and encouraged broader commercial aspirations in the New World. The short-lived War of the Quadruple Alliance [1718-1720] pitted Spain against Britain, France, Austria and the Netherlands over control of Italy. The New World territories from Texas to Florida were an important secondary theater of the conflict. France fought to extend control over Spanish lands in Texas and New Mexico from her settlement at the mouth of the Mississippi; the British in Carolina threatened Spanish Florida. While a treaty in February, 1720 concluded the hostilities in favor of the alliance, there was no major realignment of control in the Americas.\footnote{In fact the concluded peace likely heightened the prospects for South Sea Company and Royal African Company profitability, as the South Sea company expected to have the Spanish Asiento reinstated following the war - the document granting rights to supply African slaves to Spanish America. The South Sea Company had previously sourced slaves from the Royal African Company.}\footnote{See Weber (1992) p.165 ad ff.} Never-the-less, Spanish dominance in the New World was explicitly challenged, potentially raising interests and expectations among French, British and Dutch investors about future New World inroads.\footnote{See also Neal (1990). for a discussion between the integration of Dutch and British Stock Markets in the 18th century.}

The commonality in the patterns of price run-ups in Amsterdam and London suggest a high degree of financial integration between these markets.\footnote{See also Neal (1990). for a discussion between the integration of Dutch and British Stock Markets in the 18th century.} However, as noted above the West Indies Company rose later than the South Sea Company or the Royal African Company. Part of the difference can simply be attributed to time keeping: there was an eleven-day difference between the older Julian calendar used in London and the newer Gregorian calendar used on the continent. But
even accounting for this time lag, there remains at least a month difference in the
beginning of a bubble trend. This suggests that any spill-over of irrational (or
rational) exuberance ran from Britain to the Continent, not vice-versa.

2.5 International Stock Market Crash

As we have seen, share prices in several companies rose dramatically in 1720, but
the timing of their take-off differed. In this section we examine the timing of the
-crash. Figure 2.3 shows the Dutch West Indies Company, Stad Rotterdam, the
South Sea Company and the two British insurance companies. The trading dates
for the British companies have been adjusted to the Gregorian calendar. Hereafter
a G will indicate a Gregorian date, and J a Julian date.

The prices are indexed to one at the start of each series, and the vertical
axis is logarithmic to allow comparison of the scale of the price changes on a
percentage basis. The two British insurance companies reached their peak on the
same day - August 26th [G], and began their drop on the 27th [G]. They both
fell significantly over the following four trading sessions. The Dutch West Indies
Company reached its peak on August 28th [G] and began its drop on the following
trading day, August 31st. It, too, continued to fall significantly over the following
several trading sessions. Together, these three firms - aside from the Mississippi
Company in France - were the first major companies to crash in price in 1720. The
-crash in Royal African Company shares began three weeks later on September 14th
[G] and the crash in South Sea Company shares began on September 19[G]. Since
late August seems to have been the important turning part in the London and
Dutch stock markets, a natural question to ask is what happened around August
27th [G], or August 16th [J]?

The Leydse Courant reported news about both the British insurance companies
in its edition of August 28th [G] noting that, on the 23rd of August [G] the Royal
Assurance Company was planning a new issue of shares for the following week,
-presumably to raise the 50,000 pounds payment to the Exchequer promised by
September 11 [G]. This payment was a condition of their charter granted as a
result of the Bubble Act.19 The London Assurance Company was also required
to deliver the same sum on that date. The news also noted cryptically that the
London Assurance Company “kept silent and sought to learn the secrets of the
other firm” presumably alluding to some scandal.20 This suggested that there was
some asymmetric information of potential importance to investors.

Not reported in the Courant, but known from the London Gazette of August
23rd [J] is that the Attorney General issued a writ of scire facias on August 29th
[G] against four firms (including the York Building Company) seeking to expand
their business beyond their charter. This writ represented a serious risk to firms
seeking broad latitude. Although they were not named in the writ, it likely fed
negative speculation about the potential constraints to the aspirations two marine

2 New Evidence on the First Financial Bubble

Figure 2.3: Timing of the Crash

This figure displays the share price development of the Dutch West Indies Company, Rotterdam Company, South Company, Royal Exchange Assurance and London Assurance from August, 1719 to December, 1720. Prices are normalized to 1 by the first observation of the series and displayed on a logarithmic axis.
2.5 International Stock Market Crash

insurance firms to expand into fire and life insurance. Although the writ was ultimately annulled, the reprieve came after the market crash.

Non-legal factors may also have come into play around this time. The Leydse Courant of September 2 [G] reported news from London dated August 30th [G] that a fleet of twelve ships from Jamaica had been lost and they were insured by the London Assurance Company for 72,000 pounds.\footnote{The date of this event is incorrectly reported in Postlethwayt (1766) as occurring in October.} It noted the burglary of the home of one of the directors of the company.\footnote{Leydse Courant, vol. 125, p. 1, September 2nd, 1720.} The insurance claim from the fleet and possibly the burglary would have raised concerns about the capability of the firm to meet its September payment deadline.

Thus, over the course of four days in late August, 1720, adverse news about the York Building Company and the two major insurance companies hit the market and quite likely triggered their rapid decline in price. Over the course of a week, all three British firms declined roughly 30%. This decline is large enough to have caused financial distress to speculators on margin. In this way it may have led to a broader financial crisis sparked by borrowers liquidating securities to cover obligations.

Because they are reported in the Leydse Courant, we can track the dates on which these same news items reached the Netherlands. The lag of three to four days between the crash in London insurance companies and the Dutch West Indies Company is not surprising. Koudijs (2009) calculates that the average 18th century travel time of packet boats between London and Amsterdam was about three or four days. Likewise, the lag of three to four days is confirmed by evidence from the Leydse Courant’s dated bylines vs. publishing dates for London news. We might expect, however, that some reports reached the Dutch capital markets through other channels.

Thus, through public information sources, Dutch investors in the British bubble companies as well as the Dutch West Indies Company would have known by August 29th about the financial plans and troubles of the British insurance firms. Also they may also have known through private information sources that the Jamaica fleet was wrecked.\footnote{The Amsterdam insurance market would have had the same capabilities of learning about Caribbean shipwrecks in a timely manner as the London market, since these affected underwriting decisions regardless of who insured the ships.}

It is not clear whether these reports were good news or bad for Dutch investors. The West Indies Company was not engaged in the insurance trade, and the only value-relevant news about the new world (apart from the shipwrecks) might have been the August 30 [G] report in London of the discovery of gold in Jamaica, reported in the September 2 [G] Leydse Courant. The only source of propagation of the crash is through the channel of financial distress. Investors in the West Indies Company who held British insurance shares on margin might have had to raise cash by selling their Dutch assets. The Leydse Courant reported the prices of Royal Exchange Assurance and the London Assurance intermittently for the trading days July 6 [G] to August 24th [G], and for the York Building
Company from July 19th [G] to August 17th [G]. It seems reasonable to interpret this demand for news about these firms as evidence of speculative interest in these securities among Dutch investors. Interestingly enough, shares in Stad Rotterdam jumped by 15% from August 28th to August 31st. Since the firm was created to compete with the British insurance companies, the problems of a rival might have been viewed as an opportunity. Stad Rotterdam did not crash until the end of September.

The coincidence of the price peaks for three British firms and the West Indies Company seems best explained by the onset of a liquidity crisis in the international stock markets - a crisis that overwhelmed the capital markets in the following two months. The coincidence of the scire facias writ limiting British companies to their charters seems to be the leading culprit in the events that sparked the crash, although negative news about one of the British insurance firms also likely played a role.

Beyond the micro-question about what sparked the global financial crash, the one thing we can determine from the time-series of various stock prices is that, while British and Dutch firms rose at different times over the course of the first eight months of 1720, September and October were terrible months for all stocks. The crash overwhelmed all stocks regardless of whether investor expectations were based on prospects of Atlantic trade, maritime insurance or real estate deals.

One thing the rise in Dutch share prices reveals is that government re-funding could not have been the sole basis for the British bubble. No Dutch firm was launched to imitate the financial operations of the Mississippi Company or the South Sea Company. The fact that other types of firms bubbled - including the West Indies Company - suggests that the Mississippi and South Sea bubbles were not primarily due to speculation about the debt-equity swap as a financial innovation.

2.6 Cross-sectional Evidence from New Company Issues

The price bubbles in 1720 were accompanied in Britain and the Netherlands by a wave of new company issues. The Leydse Courant provides price information for many of the new firms launched in the Netherlands, however regular price quotations for the new companies in London have not survived. There is, however, some information to allow us to further examine the cross-sectional differences in the degree of exposure to investor enthusiasm. A satirical British print made in 1721, shortly following the crash, The Bubbler’s Mirror, lists a number of the well-known London issues along with the par value of the shares and the maximum percentage over par achieved during the bubble. This information is sufficient for us to further quantify the cross-sectional differences noted for the larger firms and to determine the extent to which differences observed across industries - i.e.

\[24\] Reported in Scott (1910) vol. 1 p.410.
Atlantic trade, insurance and banking carry through more generally. The data from The Bubbler’s Mirror is reported in Appendix 2.A.

Table 2.1 shows the average growth by industry. The first column includes the large firms previously studied. In the second column, we removed the Bank of England, the Million Bank, the South Sea Company, the Royal African Company, the Royal Exchange Assurance Company and the London Assurance Company from the calculations. The one exception is the inclusion of both columns of the East India Company as a basis for comparison.

Table 2.1: Maximum Percentage Price Increase of British Firms over par by Industry, 1720

This table reports the average growth in share price per industry. The first column includes all firms. In the second column, we removed the Bank of England, the Million Bank, the South Sea Company, the Royal African Company, the Royal Exchange Assurance Company and the London Assurance Company from the calculations. The final column reports the number of firms in each industry.

<table>
<thead>
<tr>
<th>Industry</th>
<th>Total</th>
<th>Total (less large firms)</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance</td>
<td>2013%</td>
<td>1717%</td>
<td>8</td>
</tr>
<tr>
<td>Real estate</td>
<td>1625%</td>
<td>1625%</td>
<td>2</td>
</tr>
<tr>
<td>Commodity</td>
<td>1208%</td>
<td>1208%</td>
<td>12</td>
</tr>
<tr>
<td>Manufacture</td>
<td>1166%</td>
<td>1166%</td>
<td>6</td>
</tr>
<tr>
<td>Atlantic</td>
<td>895%</td>
<td>948%</td>
<td>4</td>
</tr>
<tr>
<td>Marine</td>
<td>875%</td>
<td>875%</td>
<td>6</td>
</tr>
<tr>
<td>Service/Utility</td>
<td>567%</td>
<td>567%</td>
<td>3</td>
</tr>
<tr>
<td>Pacific</td>
<td>349%</td>
<td>349%</td>
<td>1</td>
</tr>
<tr>
<td>Bank/Finance</td>
<td>335%</td>
<td>500%</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>1172%</td>
<td>995%</td>
<td>45</td>
</tr>
</tbody>
</table>

In order to test the hypothesis that the British firms associated with the Atlantic trade reached higher premia over par values, we performed a t-test on the log growth rates of the Atlantic trade firms and tested the null that the growth rate for the East India Company was drawn from the same distribution. Despite the few degrees of freedom, we were able to reject the null with greater than 85% confidence for both specifications. In growth rates, the t-value was 4.25 on 3 df. For log growth the t-value was 6.15 on 3 df. Both are significant at traditional confidence levels. The important caveat is that the small sample and its unknown distributional properties potentially limits the interpretation of parametric tests.

The Table shows that insurance industry was the highest growth industry, although it also had the highest cross sectional variation in rates: Four of the top 25

The test was performed for two specifications: simple growth rates and logged growth rates.

21
seven firms ranked by growth were insurance companies, and two of the last seven were insurance companies. Column two shows that excluding the two top marine insurance companies, Royal Exchange and London Assurance reduced the scale of the bubble in insurance firms, but did not change its top rank. This is due to the fact that two firms (General Insurance and British Insurance) bubbled on a comparable scale to their more widely traded cousins. A t-test of the difference in mean growth rates between insurance companies and all other firms in the sample yielded mixed results.\footnote{A t-test assuming differences in unknown variance between the samples yielded a probability value of 22\% for logged growth rates and 0.3\% for raw growth rates.}

The evidence strongly suggests that the exuberance of London investors was driven by certain industries. While the Atlantic trade is the obvious candidate given the fame of the South Sea Company, the data suggest that there was a major bubble - perhaps even a larger bubble - in insurance. The prominent position of the insurance companies in the bubble was noted by contemporary observers; most famously, John Aislabie, Chancellor of the Exchequer who took a bribe of 20,000 pounds of South Sea stock in return for his political support of the firm. In his unsuccessful defense before the House of Lords, Aislabie exclaimed of the two insurance firms: "these two projects were founded in greater iniquity and contributed more to the publick calamity than anything else."\footnote{Mr. Aislabie’s second speech on his defense in the House of Lords, London (1721) p. 14 quoted in Scott (1910) p. 405.}

An interesting feature of the Bubble Act of June 9, 1720 which asserted governmental control over chartering companies for limited purposes is that it did not directly concern the South Sea Company. Rather, it chartered the London Assurance and the Royal Exchange Assurance companies and explicitly limited the chartering of competitors. It thus makes sense to look at the bubble in 1720 through the lens of financial innovation in the insurance sector as much as from the perspective of an event driven by massive government debt conversion, or indiscriminant speculation in company shares.\footnote{For a view of the effect of the Bubble Act on business organization, see Harris (1994) and Harris (2000).}

### 2.7 The IPO Wave in the Netherlands

In the Netherlands, a number of new firms were capitalized in 1720, beginning in July with the creation of Stad Rotterdam and extending through October of that year. We collected data from the Leydse Courant for many of these new firms. Figure 2.4 shows the major Dutch price series’ as well as a number of the additional issues over the last six months of 1720. As noted previously, the East India Company [OIC] rose only a little through the year and then dropped below its June level by the end of 1720. The West Indies Company rose quite dramatically and raised additional capital in two subscriptions in 1720 as its stock price rose. The most striking feature of the figure is that, with the exception of Middelburg, which floated both an insurance company and a commercial company, all the Dutch
2.7 The IPO Wave in the Netherlands

Figure 2.4: Dutch IPO Wave

This figure displays the share price development of Dutch shares issued in 1720 and quoted in the Leydse Courant. Prices are extracted from the Leydse Courant, denoted in nominal amounts and plotted on a logarithmic axis for the period July, 1720 to December, 1720.
projects were singularly associated with its own city. Gelderblom and Jonker (2009) point out that this reflects the history of Dutch corporate development. The East India Company resulted from a consolidation of trading companies in different cities. Gelderblom and Jonker hypothesize that Dutch cities may have anticipated a similar consolidation of the insurance trade, and thus the first move of Rotterdam might have stimulated a rush to create similar projects that would result in a share of the consolidated firm. If this were so, it would suggest that the Dutch viewed the publicly traded corporate insurance company as an important and potentially transformative financial innovation that had the power to become a dominant organization. No city wished to be left out of the potential future gains to the consolidation of the Rotterdam company into a huge national enterprise. This is consistent with our hypothesis that speculation about the new insurance company form was a major driver of the international bubble in share prices in 1720. If the expectation of consolidation motivated the flurry of Dutch public offerings, such a consolidation never materialized. The shares of Stad Rotterdam began trading in mid-July, after the peak of the South Sea Company, but a month before the peak of the London and Royal Exchange Assurance companies. The firm made a secondary offering a month later, and by that time many other companies had been launched. Gouda, Delft and Schiedam were next, and they all followed a trajectory in August and September. After a drop following their initial offerings their shares rose sharply - increasing from the pre-September lows by as much as 100% to 300%. These brief spikes may not have been comparable in scale and duration to the bubble in West India Company shares, but they were evidently perceived by some market participants as such. The Dutch bubble was memorialized in late 1720 by a famous satirical book of prints, plays and financial ephemera entitled Het Groote Tafereel der Dwaasheid - The Great Mirror of Folly. Following the entry of Gouda, Delft and Schiedam came a flood of new issues at the end of September and the beginning of October; just as the global crisis hit London and the Netherlands. Note the density of initial public offerings and transactions in the Netherlands in September in this brief interval. The Dutch new issues market lasted a brief two months. All shares traded down in November and towards the close of the month trading dried up. Share quotations are virtually lacking in December. The Leydse Courant stopped quoting share prices in January, 1721.

The financial bubble in Holland began later than the bubbles in France or Britain, and it came to a conclusion with the crash in Britain with nearly precise correlation to the fall in British insurance company shares. Except for the dramatic run up in West Indies Company shares, the scale of the Dutch bubble was also somewhat smaller. Appendix 2.B reports the maximum trough to peak percentage gain for the full sample of Dutch firms. The average was 127%; much lower than the average maximum percentage gain in the London bubble. This is consistent with the claim of Gelderblom and Jonker (2009) that the Dutch bubble was smaller in scale than the British and French bubbles.

Assessing the economic significance of the bubble in the Netherlands based upon the peak to trough range of prices may be misleading, however. The new Dutch data support the relative importance of speculation about insurance. In
Appendix 2.B we report the stated intent of the new Dutch firms launched in 1720, gathered from their founding documents. Unlike the new British firms, which varied considerably in their industrial classifications, most of the Dutch companies were established to conduct commerce and trade. However, a majority of the companies included insurance as one of their proposed lines of business. It is telling that the charter of Stad Rotterdam mentions the competitive challenge of the new British insurance companies in motivating the need to launch the firm.\footnote{Anonymous, 1720, Het Groote Tafereel de Dwaasheid, p. 25.}

The crash in the equity markets was a significant setback to this new financial technology. The success of the public insurance corporation depended on the external demand for shares and the availability of public investment capital. These dried up with the global contraction of securities markets following the crash in Great Britain. Had the global crash not occurred, public investment in marine insurance underwriting might have continued. After 1720, many of the new Dutch firms closed and returned shareholder capital (or issued financial substitutes such as life annuities). A few firms survived and prospered. The Middelburg commerce company became a major player in the Atlantic economy - including the slave trade - in the late 18th century. Stad Rotterdam also fulfilled the promise of the potential of a publicly capitalized insurance corporation. It survived and prospered to become one of the largest insurance companies in continental Europe, recently merging with the financial giant Fortis.

The survival of Stad Rotterdam, London Assurance and Royal Exchange Assurance into the modern era proved the long-term viability of incorporated insurance companies. Given that the average peak to trough price increase for most of the Dutch companies was comparatively modest, it is difficult to view investors of the time as wild speculators. The underlying rationale for investing at the time - i.e. an innovation in the financing of marine insurance underwriting, does not appear to have been unreasonable.

\subsection{Conclusion}

The cross-sectional evidence from British and Dutch firms in 1720 does not directly allow a test of whether investment in shares at the time was economically rational. Instead, it provides more information about the nature and timing of investor expectations.

The data for major British companies suggests that expectations (rational or not) about the Atlantic trade may have been an important factor. The differential between the South Sea Company and the East India Company in the British market is matched by the differential between the Dutch East and West Indies Companies. This hypothesis is supported by a test of the size of the bubble using another data set of “bubbles”. The bubble in other Atlantic-trade firms, excluding the Royal African Company and the South Sea Company, also exceeded the maximum growth in East Indies Company shares in 1720.
We have price records for nine major firms over the year 1720; two from the Netherlands and seven from Britain. Of these, the share prices for the three firms engaged in the Atlantic trade: the South Sea company, the Royal African Company and the Dutch West Indies Company were higher at the end of 1720 than at the beginning.\textsuperscript{30} Although their prices increased by multiples in the middle of the year and then dropped considerably from their highs in 1720, as a group, they experienced a permanent price improvement. Firms engaged in other industries had mixed results. This cross-sectional evidence suggests that the "bubble" may have been based upon some fundamental common factor that justified a value increase. While investor irrationality may have carried prices to many multiples of their post-crash value, the bubble speculation may have anticipated some long-term permanent effect.

The cross-sectional data also show that speculation in insurance companies was another important factor in the bubble. The chartering and incorporation of insurance companies in the early 18th century was a financial innovation. It extended the potentially valuable feature of limited liability to firms that dealt in risk. Before this transition, insurance was provided by underwriters operating alone or in syndicates that pooled capital. With the transition to corporations came broader access to public capital. The new, liquid stock markets gave the public insurance firms the capacity to increase their capital base and to diversify their risks. These may have been perceived as valuable financial innovations at the time. In summary, a view of the stock prices of more than thirty traded companies, and the price gains in 1720 of another forty additional firms provides a useful perspective on the bubbles in South Sea Company and Mississippi Company shares. While the actions and price dynamics of these two major companies have dominated the historical study of the period because of their major scale, a cross-sectional perspective suggests that the basis for speculative enthusiasm at the time may have been connected to long-term prospects for Atlantic commerce, and the recent innovation in the organizational form of insurance companies.

\textsuperscript{30} South Seas = 45% gain, Royal African = 91% gain and West Indies Company 51% gain.
### 2.A Appendix I: The British Bubble

<table>
<thead>
<tr>
<th>Company name</th>
<th>Industry</th>
<th>Lowest</th>
<th>Highest</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuring of Land</td>
<td>Service/Utility</td>
<td>1/8</td>
<td>15</td>
<td>1100%</td>
</tr>
<tr>
<td>Furnishing of Funerals</td>
<td>Service/Utility</td>
<td>2 1/2</td>
<td>15</td>
<td>500%</td>
</tr>
<tr>
<td>Liverpool Water Supply</td>
<td>Real Estate</td>
<td>10</td>
<td>20</td>
<td>2950%</td>
</tr>
<tr>
<td>York Buildings Co.</td>
<td>Real Estate</td>
<td>10</td>
<td>5</td>
<td>200%</td>
</tr>
<tr>
<td>East India Company</td>
<td>Pacific</td>
<td>100</td>
<td>40</td>
<td>300%</td>
</tr>
<tr>
<td>Improvement of Land</td>
<td>Real Estate</td>
<td>1</td>
<td>15</td>
<td>100%</td>
</tr>
<tr>
<td>East India Company</td>
<td>Real Estate</td>
<td>5</td>
<td>70</td>
<td>300%</td>
</tr>
<tr>
<td>Building or Buying ships for Freight</td>
<td>Marine</td>
<td>1/2</td>
<td>25</td>
<td>300%</td>
</tr>
<tr>
<td>Navigation of the River Douglas</td>
<td>Marine</td>
<td>5</td>
<td>250</td>
<td>400%</td>
</tr>
<tr>
<td>Grand Fishery</td>
<td>Marine</td>
<td>1/2</td>
<td>25</td>
<td>150%</td>
</tr>
<tr>
<td>Whaling Co.</td>
<td>Marine</td>
<td>10</td>
<td>250</td>
<td>2400%</td>
</tr>
<tr>
<td>Leeds Brass Mills</td>
<td>Manufacture</td>
<td>5</td>
<td>10</td>
<td>105%</td>
</tr>
<tr>
<td>Temple Brass Mills</td>
<td>Manufacture</td>
<td>10</td>
<td>250</td>
<td>2400%</td>
</tr>
<tr>
<td>Royal Lustering Co.</td>
<td>Manufacture</td>
<td>1/8</td>
<td>40</td>
<td>1150%</td>
</tr>
<tr>
<td>Water Engine</td>
<td>Manufacture</td>
<td>2 1/2</td>
<td>105</td>
<td>1100%</td>
</tr>
<tr>
<td>Royal Lustering Co.</td>
<td>Manufacture</td>
<td>1/4</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>British Frame-work Knitters Co.</td>
<td>Manufacture</td>
<td>4 1/2</td>
<td>8</td>
<td>800%</td>
</tr>
<tr>
<td>Irish Sail Cloth</td>
<td>Manufacture</td>
<td>1/8</td>
<td>175</td>
<td>3400%</td>
</tr>
<tr>
<td>Puckle's Machine Gun</td>
<td>Insurance</td>
<td>5</td>
<td>175</td>
<td>2400%</td>
</tr>
<tr>
<td>General Insurance</td>
<td>Insurance</td>
<td>10</td>
<td>250</td>
<td>2400%</td>
</tr>
<tr>
<td>London Exchange Assurance Co.</td>
<td>Insurance</td>
<td>1/8</td>
<td>3</td>
<td>200%</td>
</tr>
<tr>
<td>British Insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Continued on next page*
Table 2.2 – continued from previous page

<table>
<thead>
<tr>
<th>Company name</th>
<th>Industry</th>
<th>Lowest</th>
<th>Highest</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rose Insurance</td>
<td>Insurance</td>
<td>1/2</td>
<td>4</td>
<td>700%</td>
</tr>
<tr>
<td>Life Insurance</td>
<td>Insurance</td>
<td>1/2</td>
<td>4</td>
<td>700%</td>
</tr>
<tr>
<td>Marine Insurance</td>
<td>Insurance</td>
<td>1</td>
<td>3</td>
<td>200%</td>
</tr>
<tr>
<td>Sun Fire Office</td>
<td>Insurance</td>
<td>10</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Gold Mining Co.</td>
<td>Commodity</td>
<td>1/2</td>
<td>16</td>
<td>3100%</td>
</tr>
<tr>
<td>Welsh Copper Co.</td>
<td>Commodity</td>
<td>4 1/8</td>
<td>90</td>
<td>2082%</td>
</tr>
<tr>
<td>English Copper Co.</td>
<td>Commodity</td>
<td>5</td>
<td>105</td>
<td>2000%</td>
</tr>
<tr>
<td>Rock Salt</td>
<td>Commodity</td>
<td>1 1/4</td>
<td>15</td>
<td>1100%</td>
</tr>
<tr>
<td>Hemp and Flax</td>
<td>Commodity</td>
<td>1/8</td>
<td>1 1/2</td>
<td>1100%</td>
</tr>
<tr>
<td>Melioration of Oil</td>
<td>Commodity</td>
<td>5</td>
<td>60</td>
<td>1100%</td>
</tr>
<tr>
<td>Saltpetre</td>
<td>Commodity</td>
<td>1/8</td>
<td>1 1/2</td>
<td>1100%</td>
</tr>
<tr>
<td>Flax and Hemp Growing in PA</td>
<td>Commodity</td>
<td>2 1/2</td>
<td>28</td>
<td>1020%</td>
</tr>
<tr>
<td>Trading with Hamburgh</td>
<td>Commodity</td>
<td>15</td>
<td>120</td>
<td>700%</td>
</tr>
<tr>
<td>Drying Malt by Air</td>
<td>Commodity</td>
<td>1/8</td>
<td>1</td>
<td>700%</td>
</tr>
<tr>
<td>Supplying Coals from Newcastle</td>
<td>Commodity</td>
<td>1/4</td>
<td>1</td>
<td>300%</td>
</tr>
<tr>
<td>Holy Island Salt</td>
<td>Commodity</td>
<td>5</td>
<td>15</td>
<td>200%</td>
</tr>
<tr>
<td>Westley’s Actions</td>
<td>Bank/Finance</td>
<td>2</td>
<td>12</td>
<td>500%</td>
</tr>
<tr>
<td>Million Bank</td>
<td>Bank/Finance</td>
<td>100</td>
<td>440</td>
<td>340%</td>
</tr>
<tr>
<td>Bank of England</td>
<td>Bank/Finance</td>
<td>100</td>
<td>265</td>
<td>165%</td>
</tr>
<tr>
<td>Bahama Islands</td>
<td>Atlantic</td>
<td>3</td>
<td>40</td>
<td>1233%</td>
</tr>
<tr>
<td>South Sea Company</td>
<td>Atlantic</td>
<td>100</td>
<td>1050</td>
<td>950%</td>
</tr>
<tr>
<td>Royal African Company</td>
<td>Atlantic</td>
<td>24</td>
<td>200</td>
<td>733%</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Atlantic</td>
<td>5 1/4</td>
<td>40</td>
<td>662%</td>
</tr>
</tbody>
</table>

Based on the Bubbler’s Mirror print Means: 13 1/4 102 1172%
from Scott vol.1 page 420 Mean (par-weighted) 670%
2.B Appendix II: The Dutch Bubble

Table 2.3: Maximum Growth for Dutch Companies

This table reports the increase in share price during the South Sea Bubble for the Dutch companies for which we collected share price information from the Leydse Courant in column 2. In column 3 we report the primary goal of the company as stated in the company’s founding documents and in column 4 if the company initiated insurance business.

<table>
<thead>
<tr>
<th>Company name</th>
<th>Price increase</th>
<th>Primary Goal</th>
<th>Insurance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIC</td>
<td>647%</td>
<td>Int. trade</td>
<td>-</td>
</tr>
<tr>
<td>Middelburg Commercie</td>
<td>406%</td>
<td>Commerce</td>
<td>-</td>
</tr>
<tr>
<td>Schiedam</td>
<td>325%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Edam</td>
<td>300%</td>
<td>Commerce</td>
<td>-</td>
</tr>
<tr>
<td>Momnikendam</td>
<td>278%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Utrecht</td>
<td>275%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Delft 1st Subscription</td>
<td>118%</td>
<td>Insurance</td>
<td>Insurance</td>
</tr>
<tr>
<td>Gouda</td>
<td>106%</td>
<td>Insurance</td>
<td>Insurance</td>
</tr>
<tr>
<td>WIC 2nd Subscription</td>
<td>103%</td>
<td>Int. trade</td>
<td>-</td>
</tr>
<tr>
<td>Naarden</td>
<td>100%</td>
<td>Manufacture</td>
<td>-</td>
</tr>
<tr>
<td>Weesp</td>
<td>100%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Muiden</td>
<td>100%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Medemblik</td>
<td>100%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Rotterdam 1st Subscription</td>
<td>76%</td>
<td>Insurance</td>
<td>Insurance</td>
</tr>
<tr>
<td>Dordrecht</td>
<td>64%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Den Haag</td>
<td>50%</td>
<td>Finance</td>
<td>Insurance</td>
</tr>
<tr>
<td>Maasshuis</td>
<td>50%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Rotterdam 2nd Subscription</td>
<td>42%</td>
<td>Insurance</td>
<td>Insurance</td>
</tr>
<tr>
<td>Vlaardingen</td>
<td>40%</td>
<td>Marine</td>
<td>Insurance</td>
</tr>
<tr>
<td>OIC</td>
<td>39%</td>
<td>Int. trade</td>
<td>-</td>
</tr>
<tr>
<td>Middelburg Assurantie</td>
<td>39%</td>
<td>Insurance</td>
<td>Insurance</td>
</tr>
<tr>
<td>WIC 3rd Subscription</td>
<td>33%</td>
<td>Int. trade</td>
<td>-</td>
</tr>
<tr>
<td>Alkmaar</td>
<td>33%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Zwolle</td>
<td>33%</td>
<td>Marine</td>
<td>Insurance</td>
</tr>
<tr>
<td>Hoorn</td>
<td>0%</td>
<td>Commerce</td>
<td>-</td>
</tr>
<tr>
<td>Veere</td>
<td>0%</td>
<td>Commerce</td>
<td>-</td>
</tr>
<tr>
<td>Brielle</td>
<td>-7%</td>
<td>Commerce</td>
<td>-</td>
</tr>
<tr>
<td>Enkhuizen</td>
<td>-10%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Purmerend</td>
<td>-17%</td>
<td>Commerce</td>
<td>Insurance</td>
</tr>
<tr>
<td>Mean</td>
<td>127%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

31 Gelderblom and Jonker (2009) note that the West Indies Company had applied to offer insurance.
2 New Evidence on the First Financial Bubble

2.C Appendix III: Price Quote Interpretation

One of the problems in the analysis of the price data from Holland in 1720 is the question of how the price quotations should be interpreted. Share or subscription prices were not quoted in currency, but in a percentage in excess of some value. Scholars to date have disagreed on the interpretation of this value. We address this basic question through examination of the language used in the price lists in 1720, and by matching share loan transactions in the books of one company to the documented market quotations. The benefit of matching price quotes to company transactions in their shares is that the books were kept in units of currency.

Price quotes in the Leydse Courant typically were given in sentences such as: "Rotterdam: Today the shares of our company were traded for prices ranging from 52 to 56 percent avans". Intervals like this are not uncommon due to the high daily volatility; especially in the rise of the bubble the morning prices differed substantially from evening quotes. A key question posed by these quotes is of course, what is meant by "avans"? What amount is the quotation in excess of? This question is more challenging than it appears. Shares were issued through subscriptions that required capital calls. Thus, the relevant multiplicand might be interpreted as the capital paid in to the company up to that date [the paid in capital] or it might be interpreted as the face value of the share after all the shareholder installments are paid – something we now often refer to as the par value or nominal share value.

It is therefore not surprising that previous authors have debated the interpretation of these quotes. For example:

Van Rijn (1899) presumed that the avans referred to the nominal (or fully paid in) value of the share, and that additional paid-in capital represented a fraction that incremented the quoted price.32

This is logical because newspaper percentages could not practically be based on the amounts paid-in. The paid-in capital increased over time and the exact amount paid-in was therefore uncertain until the books were closed and the paid-in amounts computed. If newspaper percentages were based on the amount paid-in, the newspaper percentages would have to drop after the payment of an installment; if the market share price does not change and more capital is paid-in, the percentage decreases by construction.

Smith (1919) concurs with this interpretation: "The share price was recorded in percentages "avans", i.e. it indicated the percentages to be paid in excess of par value, calculated based on the amount which, with respect to the nominal amount

32 "The furnishment up till 80% of the shares of the first subscription is also foolish and absurd, since the amount furnished in these shares exceeds the market value of the original shares, i.e. 15% has been paid in on the subscription shares, to demand an additional 65% would be highly unfair ... The shares of the first subscription have already been traded at 38%, together with the paid-in amount of 15% this makes 53 percent, if an additional 65% were to be paid in then these shares would be worth 118%, this would be unreasonable. ... If ... on the old shares 75% had been paid in, these would now be worth 3750 guilders, so a share of 1000 guilders would be worth 750 guilders."
traded by the company as installment was demanded." 33 Likewise, Slechte (1982): "In the first two months the prices fluctuated between 100 and 80%, but dropped in September and October to 60%. The Rotterdam shares were therefore on average priced at 4000 guilders in this period."

Despite their general agreement, none of the earlier authors offered empirical evidence supporting their interpretations. Thus, we sought to reject or verify their calculations of monetary share prices by matching transactions in the Leydse Courant with company records.

After the burst of the bubble at the end of September 1720, the directors of Stad Rotterdam made non-recourse loans to directors using their shares as collateral. If share prices fell during the maturity of the loan, the company incurred the loss. This arrangement could be interpreted as beneficial to shareholders because it was a means to reduce the "float" of shares by keeping directors from dumping into a falling market. It could also be interpreted as pure self-dealing by directors. In either case, these transactions allow us to link the market quotes to currency-valued transactions.

On pages 37 and 142 of the general ledger of Stad Rotterdam, some of these share loans are recorded in prices that match quotations reported in the Leydse Courant for the same day. 34 Since the ledger accounts are recorded in guilders, we can ascertain that the newspaper prices are reported in percentages of the nominal share value (5000 guilders for the Rotterdam company). The losses incurred in these loan transactions also appear on the profit and loss account of the Rotterdam company (page 9 of the general ledger account).

A simple example of our interpretation is useful. Suppose that Rotterdam shares [with 5,000 guilders nominal value] are trading for 2,500 guilders on a certain date and that up until that date 1% had been paid in. Then the Leydse Courant would quote of 49% avans which would imply the following: "Rotterdam is trading today at 50% = (2,500/5000 * 100%) or 49% of the par value of the share above the paid-in capital = 49% = ((2,500-1%*5,000)/5,000)*100%".

This convention was convenient during weeks of installment payments. It allowed investors to distinguish between shares on which the installment has been paid and shares on which the installment still needed to be paid.

33 Original Dutch text: "De koers werd genoteerd in procenten "avans", dus men gaf aan hoeveel procenten boven pari betaald moest worden, te berekenen over de geldsom, die, met betrekking tot het verhandelde nominale bedrag, door de compagnie als storting was geëist."

34 Gemeentearchief Rotterdam, archive no. 199 inventory no. 451, pages 37 and 142.
Chapter 3

Efficient Estimation of Firm-Specific Betas and its Benefits for Asset Pricing Tests and Portfolio Choice

We improve both the specification and estimation of firm-specific betas. Time variation in betas is modeled by combining a parametric specification based on economic theory with a non-parametric approach based on data-driven filters. We increase the precision of individual beta estimates by setting up a hierarchical Bayesian panel data model that imposes a common structure on parameters. We show that these accurate beta estimates lead to a large increase in the cross-sectional explanatory power of the conditional CAPM. Using the betas to forecast the covariance matrix of returns also results in a significant improvement in the out-of-sample performance of minimum variance portfolios.1

3.1 Introduction

Precise estimates of firm-specific betas are crucial in many applications of modern finance theory, including asset pricing, corporate cost-of-capital calculations, and risk management. For instance, portfolio managers often have to ensure that their risk exposure stays within predetermined limits and managers need estimates of their company’s beta to make capital budgeting decisions. Academics and practitioners have taken two approaches to estimating betas. Under the first one stocks are grouped into portfolios to reduce measurement error, assuming that all stocks within a given portfolio share the same beta (e.g., Fama and MacBeth (1973)). The

1 This chapter is based on the paper Cosemans, Frehen, Schotman, and Bauer (2009)
second method consists of estimating separate time series regressions for each firm to obtain individual betas (e.g., Brennan, Chordia, and Subrahmanyan (1998)).

Apart from this lack of consensus in the literature about the best method to estimate betas, existing studies also fail to provide clear guidance on the best way to model betas. Although a large body of empirical evidence suggests that betas vary over time, existing work uses different specifications to model these changes in betas. Many studies use a parametric approach proposed by Shanken (1990), in which variation in betas is modeled as a linear function of conditioning variables. An alternative, non-parametric approach to model risk dynamics is based on purely data-driven filters, including short-window regressions (Lewellen and Nagel (2006)) and rolling regressions (Fama and French (1997)).

In this chapter we improve both the specification and estimation of firm-specific, time-varying betas. We improve the specification of betas by combining the parametric and non-parametric approaches to modeling time variation in betas. Because the key strengths of each approach are the most important weaknesses of the other, we argue and show that a combination of the two methods leads to more accurate betas than those obtained from each of the two approaches separately. We allow the optimal mix of the two methods to vary across stocks, since individual firms may benefit more or less from either specification, and over time, because the preferred combination during stable market conditions may be different from that in turbulent time periods.

The parametric specification is appealing from a theoretical perspective because it explicitly links time variation in betas to macroeconomic state variables and firm characteristics (e.g., Gomes, Kogan, and Zhang (2003)). However, the main drawback of this approach is that the investor’s set of conditioning information is unobservable. Ghysels (1998) shows that misspecifying beta risk may result in serious pricing errors that might even be larger than those produced by an unconditional asset pricing model. In addition, this method can produce excessive variation in betas due to sudden spikes in the macroeconomic variables that are often used as instruments. Finally, many parameters need to be estimated when a large number of conditioning variables is included, which leads to noisy estimates when applied to stocks with a limited number of time series observations. An important advantage of the non-parametric approaches is that they preclude the need to specify conditioning variables, which makes them more robust to misspecification. However, the time series of betas produced by a data-driven approach will always lag the true variation in beta, because using a window of past returns to estimate the beta at a given point in time gives an estimate of the average beta during this time period. Although reducing the length of the window results in timelier betas, the estimation precision of these betas will also...

---


3 An alternative approach has been proposed by Christoffersen, Jacobs, and Vainberg (2009), who calculate forward-looking betas using the information embedded in option data. A drawback of this method is that it requires a cross-section of liquid stock options, which is not available for many small firms.
decrease.

We improve the estimation of individual stock betas by setting up a Bayesian panel data model that exploits the information in the cross-section of firms to obtain more precise estimates. In particular, we specify hierarchical prior distributions that impose a common structure on parameters while still allowing for cross-sectional heterogeneity. Bayesian methods are especially attractive in settings with individual-level heterogeneity in multiple parameters, because only the parameters of the hierarchical priors where the parameters are assumed to be drawn from have to be estimated. In contrast, methods that estimate every parameter individually without linking it to similar parameters, such as estimating separate time series regressions for every single firm, suffer from poor estimation precision, particularly when the number of time series observations is limited. Intuitively, the Bayes estimator can be interpreted as a weighted average of the least squares estimator for a given cross-section unit and the cross-sectional average coefficient. The Bayes estimator of the firm-specific parameters shrinks the least squares estimator towards the cross-sectional mean. When the number of observations increases, the weight gradually shifts from the prior to the data.

Our panel data approach uses both daily returns and monthly firm-level characteristics to capture the cross-sectional heterogeneity and time series dynamics in monthly betas. Including cross-sectional information increases the accuracy of firm-specific betas because previous studies document a strong cross-sectional relationship between beta and firm characteristics (see, e.g., Fama and French (1992)). Existing work further shows that the use of high-frequency returns yields more precise and timelier estimates of beta than using monthly returns (see, e.g., Bollerslev and Zhang (2003)). We use daily returns instead of intraday returns because market microstructure frictions put an upper limit on the frequency that can be used to estimate betas in practice. We combine the data sampled at different frequencies by implementing the mixed data sampling (MIDAS) approach of Ghysels, Santa-Clara, and Valkanov (2005), which determines the optimal weights given to past data.

We estimate the model using a large panel of individual stocks, which offers several advantages over the alternative of aggregating stocks into portfolios based on characteristics. First, aggregating stocks into portfolios may conceal important information contained in individual stock betas. Ang, Liu, and Schwarz (2008) show that risk premia can be estimated more precisely using individual stocks instead of portfolios, because creating portfolios reduces the cross-sectional variation in betas. A second important drawback is that due to the strong factor structure in the 25 size-B/M sorted portfolios that are often used as test assets in asset pricing studies, traditional cross-sectional tests are flawed and have low power to reject a model, as shown by Lewellen, Nagel, and Shanken (2008). Third, when stocks are grouped into portfolios based on characteristics that have been identified by previous research as determinants of average returns instead of being based on economic theory, the evidence against asset pricing models may be overstated because of data-snooping biases (Lo and MacKinlay (1990)).

Despite the benefits of using individual stocks, most asset pricing studies use
characteristics sorted portfolios because it is difficult to estimate firm-level parameters with a reasonable degree of precision when the number of observations is limited. Notable exceptions are Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006a), who use a two stage approach to study the impact of characteristics on risk-adjusted returns. However, both studies estimate separate time series regressions for every single firm, which leads to imprecise beta estimates, particularly for firms with a short return history.

Fama and French (2008) even conclude that “given the imprecision of beta estimates for individual stocks, little is lost in omitting them from the cross-section regressions”.

Our main empirical findings are as follows. First, we show that modeling time-varying betas as a function of both conditioning variables and past returns dominates traditional specifications in which betas depend on only one of those components. Combining these specifications produces superior beta estimates because they capture different aspects of beta dynamics. We also find that the optimal mix of these specifications varies both over time and across stocks.

Second, we show that our panel data approach produces more accurate estimates of firm-specific betas than those obtained from the traditional approach of estimating a separate time series regression for every firm. Specifically, for the average firm the posterior standard deviation of beta is significantly larger in time series regressions than in the panel model. Third, we document strong cross-sectional heterogeneity in firm betas within the 25 size-B/M portfolios that are commonly used to test asset pricing models. This confirms that aggregating stocks into portfolios conceals important information contained in individual stocks and shrinks the cross-sectional variation in betas.

We demonstrate that a more precise estimation and better specification of firm betas has important benefits for asset pricing tests and portfolio choice. In particular, we show that the betas generated by our model have significant explanatory power for the cross-section of returns. Using stocks as test assets and estimating betas in a panel model results in more efficient parameter estimates in cross-sectional asset pricing tests than using portfolios. The estimate of the market premium is positive and statistically significant, even after controlling for firm characteristics. We illustrate the value of our beta specification and estimation method for portfolio choice by using the betas to forecast the covariance matrix of stock returns. We find that the global minimum variance portfolio that is formed using this covariance matrix outperforms minimum variance portfolios based on other strategies, including the naive 1/N rule, the sample covariance matrix, and a static factor model for estimating covariances.

The chapter proceeds as follows. In section 3.2 we introduce our specification for time-varying betas in a panel data framework. Section 3.3 explains the Bayesian approach to inference and section 3.4 describes the data. We report our empirical results in section 3.5 and discuss the asset pricing and portfolio choice applications in section 3.6. Section 3.7 concludes.
3.2 The Model

In this section we describe our model for individual betas. For simplicity, we discuss our approach in a conditional CAPM setting but it is straightforward to extend our work to multifactor models.

Our goal is to show how to improve the specification and estimation of firm-specific, time-varying betas in any factor model. We start from the following panel data model for excess returns on individual stocks,

\[ r_{it} = \alpha_i + \beta_{it-1}r_{Mt} + \epsilon_{it}, \]  

(3.1)

where \( r_{it} \) is the excess return on stock \( i \) in month \( t \), \( \alpha_i \) is the risk-adjusted return, \( \beta_{it-1} \) is the conditional market beta, \( r_{Mt} \) is the excess market return, and \( \epsilon_{it} \) is a zero-mean, normally distributed idiosyncratic return shock. Following Avramov and Chordia (2006b), we assume that the covariance matrix of these shocks is diagonal and that idiosyncratic volatility is constant.

Our specification for the conditional beta consists of two components: one part is the realized beta, \( b_{it} \), and the other part is the fundamental beta, \( \beta^*_it \),

\[ \beta_{it} = \phi_{it}b_{it} + (1 - \phi_{it})\beta^*_{it}, \]  

(3.2)

where \( \phi_{it} \) and \( (1 - \phi_{it}) \) measure the proportion of the beta of firm \( i \) that is explained by the realized beta and fundamental beta, respectively. Hereafter we refer to this mixture of realized and fundamental betas as the mixed beta. We allow the optimal combination of fundamental and realized betas to vary not only across firms but also over time. Time variation in \( \phi_{it} \) is modeled as a linear function of market volatility, because the best mix of fundamental and realized betas in turbulent market conditions can be very different from that in stable periods,

\[ \phi_{it} = \phi_0 + \phi_1V_{Mt}, \]  

(3.3)

where \( V_{Mt} \) is the realized market variance, which we calculate by summing the squared daily market returns over the past year. We take the logarithm of the market variance to reduce the impact of outliers and then subtract its time series average and divide by its standard deviation, so that it has mean zero and standard deviation equal to one.

\( b_{it} \) is the realized beta that we estimate using daily data according to the Mixed Data Sampling (MIDAS) approach introduced by Ghysels, Santa-Clara, and Valkanov (2005).

We choose to estimate realized betas using daily returns because these provide a reasonable balance between efficiency and robustness to microstructure noise (see, Campbell, Lo, and MacKinlay (1997)). However, even at a daily frequency the betas of less liquid stocks might be biased downward. Following Scholes and Williams (1977), we therefore control for nonsynchronous trading effects by adding
the covariance of the stock’s return with the one-day lagged market return,\(^4\)

The MIDAS approach differs from traditional rolling window estimators of betas by selecting the optimal window for estimating betas using a flexible weighting function. Ghysels, Santa-Clara, and Valkanov (2005) use the MIDAS approach to estimate the market’s conditional variance and find that it is superior to traditional GARCH and rolling window methods.

In particular, our MIDAS estimator of realized betas is given by:

\[
b_{it} = \frac{\sum_{\tau=1}^{\tau_{\text{max}}} w_{t-\tau} r_{it-\tau} r_{Mt-\tau}}{\sum_{\tau=1}^{\tau_{\text{max}}} w_{t-\tau} r_{Mt-\tau}}, \tag{3.4}
\]

where \(t\) refers to a particular month, \(\tau\) to a particular trading day, and \(w_{t-\tau}\) to the weight given to the product of the return on stock \(i\) and the market return, \(r_{Mt-\tau}\), and to the squared market return, \(r_{Mt-\tau}^2\), on day \(t - \tau\). We set the maximum window length \(\tau_{\text{max}}\) equal to 250 days, which is approximately one year of trading days.

We parameterize the weights as a beta function:

\[
w_{t-\tau} = \frac{f\left(\tau; \kappa_1, \kappa_2\right)}{\sum_{\tau=1}^{\tau_{\text{max}}} f\left(\tau; \kappa_1, \kappa_2\right)}, \tag{3.5}
\]

where \(f\left(\tau; \kappa_1, \kappa_2\right)\) is the density of a beta distribution. As pointed out by Ghysels, Santa-Clara, and Valkanov (2005), the specification based on the beta function has several advantages.

First, it ensures that the weights are positive and sum to one. Second, it is parsimonious because only two parameters need to be estimated. Third, it is flexible as it can take various shapes for different values of the two parameters. We impose a downward sloping pattern on the weights by setting \(\kappa_1\) equal to 1, which further reduces the number of parameters that need to be estimated. \(\kappa_1 = \kappa_2 = 1\) implies equal weights, which corresponds to a rolling window estimator of beta on daily data. \(\kappa_1 = 1\) and \(\kappa_2 > 1\) correspond to the case of decaying weights. In general, the higher \(\kappa_2\), the faster the rate of decay and the quicker beta responds to new information.

\(\beta^*_t\) is the fundamental beta, parameterized as a function of conditioning variables,

\[
\beta^*_t = \delta_0 + \delta'_1 [Z_{it} \otimes BC_t], \tag{3.6}
\]

where \(Z_{it}\) is a vector that contains \(L\) firm characteristics and \(BC_t\) is a vector that contains a constant and \(M\) business cycle variables. This specification allows the relation between beta and firm characteristics to vary over the business cycle. Modeling beta dynamics as a linear function of a set of predetermined instruments

---

\(^4\) Scholes and Williams (1977) also include a lead term to capture the impact of non-synchronous trading on the market return. We only include a lag term because otherwise the model cannot be used to forecast betas.
goes back to Shanken (1990) and is consistent with the economic motivation for conditional asset pricing models, in which the stochastic discount factor is a function of macroeconomic state variables and factor premia.

We include both firm-specific and macroeconomic variables as instruments for fundamental betas because of their documented predictive power for returns (Fama and French (1989) and Lewellen (1999)). Empirical evidence that systematic risk is related to firm characteristics and business cycle variables is provided by, among others, Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Avramov and Chordia (2006a), and Goetzmann, Watanabe, and Watanabe (2008). The theoretical motivation for choosing firm characteristics as instruments is given by Berk, Green, and Naik (1999) and Gomes, Kogan, and Zhang (2003), who show that the ability of size and book-to-market to explain the cross-section of returns is due to their correlation with the true conditional market beta. They decompose firm value into the value of assets in place and the value of growth options and demonstrate that size captures the component of a firm’s systematic risk related to its growth options whereas the book-to-market ratio is a measure of the risk of the firm’s assets in place. Zhang (2005) extends this work and argues that because of costly reversibility of capital value firms have countercyclical betas while betas of growth stocks are procyclical. Because the price of risk is also countercyclical his model can explain the value premium within a rational framework. In addition to size and B/M, we also select firm-specific momentum as a conditioning variable to examine whether the momentum effect is related to beta dynamics. Theoretical support for including macroeconomic variables is provided by Santos and Veronesi (2004), who show within a general equilibrium model that market betas vary substantially with the business cycle.

Our choice of business cycle variables is motivated by previous work (e.g., Ferson and Harvey (1999)) and includes the default spread, dividend yield, one-month T-bill rate, and term spread.

Substituting equations (3.2) and (3.6) into equation (3.1) leads to the following specification:

\[ r_{it} = \alpha_i + \phi_{it}b_{it-1}r_{Mt} + (1 - \phi_{it})(\delta_0 + \delta_1[Z_{it-1} \otimes BC_{t-1}])r_{Mt} + \epsilon_{it}. \]  

A key objective in this chapter is to determine whether the time series dynamics and cross-sectional variation in betas is better described by lagged firm characteristics and macroeconomic state variables, by past realized betas, or by a linear combination of both. Therefore, we are primarily interested in the parameter \( \phi_{it} \) and compare three different specifications: (1) mixed beta (\( \phi_{it} \) unrestricted) (2) fundamental beta (\( \phi_{it} = 0 \)) (3) realized beta (\( \phi_{it} = 1 \)).
3 Efficient Estimation of Firm-Specific Betas

3.3 Methodology

3.3.1 Bayesian Methods

We estimate the model parameters using Bayesian methods.\(^5\) The main advantage of Bayesian inference in our setting is that it allows a very flexible specification for describing the dynamics in beta by imposing a common structure on the model parameters.

Updating beliefs according to Bayes’ theorem implies that the joint posterior density of the parameters, \(p(\theta|y)\), is proportional to the likelihood times the prior density.

\[
p(\theta|y) \propto p(y|\theta)p(\theta)
\]  

(3.8)

where \(\theta = (\alpha, \phi_0, \phi_1, \delta_0, \delta_1, \sigma^2_{\epsilon}, \sigma^2_{\alpha}, \sigma^2_{\phi_0}, \sigma^2_{\phi_1}, \Omega_{\delta_0}^{-1}, \Omega_{\delta_1}^{-1})\), is the set of all parameters and \(y\) is the full set of data.

The likelihood function for the model in equation (3.7) is given by

\[
p(y|\theta) = \prod_{i=1}^{N} \prod_{t \in T_i} (\sigma^2_{\epsilon_i})^{-\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma^2_{\epsilon_i}} (r_{it} - \alpha_i - r_{Mt} \beta_{it-1})^2 \right]
\]  

(3.9)

where \(\sigma^2_{\epsilon_i}\) is the idiosyncratic return variance, \(\beta_{it-1}\) is defined as in equation (3.2), \(N\) is the number of stocks, and \(T_i\) is the number of monthly return observations for firm \(i\).

3.3.2 Prior Distributions

We specify conditionally conjugate, independent, hierarchical priors that impose a common structure on the model parameters while still allowing parameters to vary across firms. Thus, our setup combines the benefits of a portfolio approach to estimating betas (e.g., Fama and MacBeth (1973)) and an approach in which separate regressions are estimated for each firm (e.g., Avramov and Chordia (2006a)). Specifically, our choice of prior distributions is as follows:

\(^5\) Bayesian methods have been used in a number of asset pricing studies, including Shanken (1987), Harvey and Zhou (1990), Kandel, McCulloch, and Stambaugh (1995), and Cremers (2006). All these studies focus on portfolios and assume that betas are constant. Ang and Chen (2007) and Jostova and Philipov (2005) use Bayesian techniques to obtain time-varying portfolio betas, which they model as latent autoregressive processes.
3.3 Methodology

\begin{align*}
\alpha_i & \sim N(0, \sigma^2_\alpha) \quad \text{with} \quad \sigma^2_\alpha & \sim IG(0.001, 0.001), \\
\phi_0 & \sim N(0.5, \sigma^2_\phi_0) \quad \text{with} \quad \sigma^2_\phi_0 & \sim IG(0.001, 0.001), \\
\phi_1 & \sim N(0, \sigma^2_\phi_1) \quad \text{with} \quad \sigma^2_\phi_1 & \sim IG(0.001, 0.001), \\
\delta_0 & \sim N(0, \sigma^2_\delta_0) \quad \text{with} \quad \sigma^2_\delta_0 & \sim IG(0.001, 0.001), \\
\delta_1 & \sim N(0, \Omega_\delta_1) \quad \text{with} \quad \Omega_\delta_1^{-1} & \sim Wish \left( [(L + LM)I]^{-1}, (L + LM) \right), \\
\sigma^2_{\epsilon_i} & \sim IG(0.001, 0.001). 
\end{align*}

We use diffuse priors to minimize their influence on the posterior densities. Following Jostova and Philipov (2005), we specify noninformative prior distributions for the variance parameters \(\sigma^2_\alpha\), \(\sigma^2_\phi_0\), \(\sigma^2_\phi_1\) and the idiosyncratic variance \(\sigma^2_{\epsilon_i}\), by setting the scale and shape parameters \(A\) and \(B\) of their inverse gamma (IG) prior distributions equal to 0.001. We set the degrees of freedom parameter \(\psi\) of the Wishart prior for \(\Omega_\delta_1^{-1}\) equal to the dimension of this matrix, \((L + LM)\), because this value gives the lowest possible weight to the prior information (see Gelman, Carlin, Stern, and Rubin (2004)). We set the scale matrix of the Wishart prior equal to \([(L + LM)I]^{-1}\), so that the prior mean of \(\Omega_\delta_1^{-1}\) is equal to the identity matrix. We give equal prior weight to the fundamental beta and the realized beta by setting the prior mean of \(\phi_0\), i.e. \(\mu_{\phi_0}\), equal to 0.5 and the prior mean of \(\phi_1\) equal to 0.6.

We parameterize the MIDAS weights as a beta function and set \(\kappa_1\) equal to 1. To rule out cases where more recent returns receive less weight than observations in the more distant past, i.e., when \(\kappa_2 < 1\), we constrain \(\kappa_2\) to the interval \([1,26]\). When \(\kappa_2 = 1\) all 250 days receive equal weight in the estimation and when \(\kappa_2 = 26\) the cumulative weight given to the 40 most recent days is 99\%. We implement this restriction by a change of variable, \(\kappa_2 = 1 + 25\kappa_2^*\).

For \(\kappa_2^*\) we choose a uniform prior, \(\kappa_2^* \sim U[0,1]\).

3.3.3 Bayesian Inference

We employ Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior distribution of the parameter vector \(\theta\). The main idea is to construct a Markov chain such that the chain converges to a unique stationary distribution that is the posterior density, \(p(\theta|y)\). We use the Gibbs sampler, which involves the sequential drawing from the full conditional posterior densities, to obtain draws from the joint posterior density. In particular, first the parameter vector \(\theta\) is partitioned into \(B\) blocks \((\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)})\). At each iteration of the Gibbs sampler each block is sampled from its posterior distribution conditional on all other blocks and the data. Because the conditional posterior density of \(\kappa_2\) has a nonstandard form, we cannot directly sample from it. Therefore, we use the Metropolis-Hastings algorithm, in which candidate parameter values are drawn

\footnote{We also considered specifications with \(\mu_{\phi_0}\) set equal to 0 or 1. These results are available upon request and show that our findings are robust to the choice of this prior distribution.}
from a proposal density and accepted with a certain probability that is highest in
areas of the parameter space where the posterior density is highest (see Chib and
Greenberg (1995)). Details on the derivation of the joint posterior density and the
conditional posterior distributions are provided in the appendix.

Iterations of the chain converge to draws from the joint posterior. We check
convergence by inspecting the standardized cumsum statistics, as suggested by
Bauwens, Lubrano, and Richard (1999), applying the partial means test based
on numerical standard errors, explained by Geweke (2005), and calculating the
Gelman-Rubin statistic that compares the variation in output between and within
chains, described by Gelman, Carlin, Stern, and Rubin (2004). These diagnostics
indicate that the parameter chains have converged after 1,000 iterations. In our
empirical analysis we therefore run 5,000 iterations and discard the first 1,000
iterations as burn-in period.

The remaining draws are used to summarize the posterior density and to con-
duct inference.

3.4 Data

The firm data comes from CRSP and Compustat and consists of the monthly
return, size, and book-to-market value for a sample of NYSE- and AMEX-listed
stocks. To calculate realized betas we further retrieve daily returns from CRSP
for these stocks. The sample covers the period from July 1964 to December 2006.
Following Avramov and Chordia (2006a), we include a stock in the analysis for a
given month \( t \) if it satisfies the following criteria. First, its return in the current
month \( t \) and in the previous 36 months has to be available. Second, data should
be available in month \( t-1 \) for size as measured by market capitalization and for
the book-to-market ratio. We calculate the book-to-market ratio using accounting
data from Compustat as of December of the previous year. Finally, in line with
Fama and French (1993), we exclude firms with negative book-to-market equity.
Imposing these restrictions leaves a total 5,017 stocks over the full sample period
and an average of 1,815 stocks per month.

Table 3.1 presents summary statistics for the data set. Panel A reports the
mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values
of excess stock returns and firm characteristics across all data points. The average
monthly excess stock return is 0.69% while the median is -0.16%. The mean (me-
dian) firm size is $1.59 (0.16) billion. Because the book-to-market ratio contains
some extreme values, we trim all book-to-market outliers to the 0.5th and 99.5th
percentile values of the distribution. After trimming, the average (median) book-
to-market ratio equals 0.96 (0.75). The cumulative return over the twelve months
prior to the current month, which we use as a proxy for momentum, has a mean of
14.65% and a median of 8.60%. Because the distributions of firm size and book-
to-market display considerable skewness, we use the logarithmic transformations
of these variables in the analysis. Furthermore, we normalize the characteristics
by expressing them as deviations from their cross-sectional means to remove any
3.4 Data

We further retrieve data for the four macroeconomic variables that we use as instruments for the fundamental beta, i.e., the default spread, dividend yield, one-month Treasury bill rate, and term spread. We define the default spread as the yield differential between bonds rated BAA by Moody’s and bonds with a Moody’s rating of AAA. The dividend yield is calculated as the sum of the dividends paid on the value-weighted CRSP index over the previous 12 months divided by the current level of the index. The term spread is defined as the yield difference between ten-year and one-year Treasury bonds. Panel B shows descriptive statistics for the macroeconomic variables. The average default spread is 1.02%, the mean dividend yield equals 3.01%, the average one-month T-bill rate is 5.69%, and the average term spread is 0.85%.

Table 3.1: Summary Statistics for Firm Characteristics and Macroeconomic Variables

This table presents descriptive statistics for stock returns, firm characteristics, and macroeconomic variables for 510 months from July 1964 through December 2006. Panel A reports the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of firm characteristics for a total of 5,017 stocks over the full sample period and an average of 1,815 stocks per month. We include a stock in the sample for a given month if it satisfies the following criteria. First, its return in the current month, $t$, and over the past 36 months has to be available. Second, data should be available in month $t-1$ for size as measured by market capitalization and for the book-to-market ratio. We exclude firms with negative book-to-market equity. XRET is the return in excess of the risk-free rate, MV represents the market capitalization in billions of dollars, and BM is the book-to-market ratio, for which values smaller than the 0.5th percentile and values greater than the 99.5th percentile are set equal to the 0.5th percentile and 99.5th percentile values, respectively. MOM is the cumulative return over the twelve months prior to the current month. Panel B shows the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of macroeconomic variables. DEF is the default spread, defined as the yield differential between bonds rated BAA by Moody’s and bonds with a Moody’s rating of AAA. DY is the dividend yield on the value-weighted CRSP index. The dividend yield is calculated as the sum of the dividends paid on the index in the previous year divided by the current level of the index. TBILL is the one-month Treasury bill rate. TERM is the term spread, defined as the yield difference between ten-year and one-year Treasury bonds.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>5th</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Firm characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XRET (%)</td>
<td>0.69</td>
<td>12.31</td>
<td>-17.48</td>
<td>-5.99</td>
<td>-0.16</td>
<td>6.42</td>
<td>21.29</td>
</tr>
<tr>
<td>MV ($ billions)</td>
<td>1.59</td>
<td>5.60</td>
<td>0.01</td>
<td>0.03</td>
<td>0.16</td>
<td>0.81</td>
<td>6.67</td>
</tr>
<tr>
<td>BM</td>
<td>0.96</td>
<td>0.82</td>
<td>0.18</td>
<td>0.44</td>
<td>0.75</td>
<td>1.22</td>
<td>2.45</td>
</tr>
<tr>
<td>MOM (%)</td>
<td>14.65</td>
<td>49.29</td>
<td>-49.12</td>
<td>-14.24</td>
<td>8.60</td>
<td>34.46</td>
<td>96.30</td>
</tr>
</tbody>
</table>

| **Panel B: Macroeconomic variables** |       |           |       |       |        |       |       |
| DEF (%)              | 1.02  | 0.43      | 0.55  | 0.73  | 0.90   | 1.21  | 1.92  |
| DY (%)               | 3.01  | 1.10      | 1.30  | 2.02  | 2.96   | 3.77  | 4.84  |
| TBILL (%)            | 5.69  | 2.70      | 1.56  | 4.08  | 5.16   | 6.96  | 10.57 |
| TERM (%)             | 0.85  | 1.14      | -1.14| 0.08  | 0.78   | 1.69  | 2.83  |
3 Efficient Estimation of Firm-Specific Betas

3.5 Empirical Results

In section 3.5.1 we study whether betas are driven by lagged conditioning variables or past realized betas. Section 3.5.2 compares the efficiency of the beta estimates produced by the panel model to that of those obtained from time series regressions. In section 3.5.3 we illustrate the loss of information from aggregating stocks into portfolios by showing the cross-sectional variation in firm-level betas within the 25 size-B/M portfolios that are often used to test asset pricing models.

3.5.1 Beta Specification

A key objective in this chapter is to improve the specification of time-varying betas. We investigate whether the time series and cross-sectional variation in betas is best explained by lagged firm characteristics and macroeconomic variables, by past realized betas, or by a linear combination of both. We address this question by estimating the model in equation (3.7) and examining the distribution of $ϕ_{it}$, which measures the proportion of beta explained by past realized beta. We first calculate $ϕ_{it}$ based on equation (3.3) for each draw of the Gibbs sampler. We then calculate for each firm the time series average $ϕ_i$ and its posterior mean.

Figure 3.1 shows the cross-sectional distribution of these posterior means of $ϕ_i$. The cross-sectional average is 0.51, which implies that for the average firm the estimate of beta is the average of the fundamental and realized beta estimates. The spread in the distribution shows that for some firms past realized betas are more important determinants of mixed betas while for others lagged fundamental betas have a stronger impact.

Figure 3.2 plots the evolution of the cross-sectional average of $ϕ$ through time. Interestingly, $ϕ$ increases during periods of high market volatility, such as recessions and the stock market crash in 1987. This implies that more weight should be given to past realized betas and less weight to fundamental betas during turbulent conditions. Because the fundamental beta specification is a function of macroeconomic and firm-specific variables, it captures long-run movements in beta driven by structural changes in the economic environment and in firm- or industry-specific conditions. In contrast, because the realized beta specification is based on high-frequency returns, it picks up short-run fluctuations in beta in periods of high market volatility.\(^7\)

Since we find that conditioning variables motivated by economic theory are important determinants of beta, we now consider the posterior distributions of the parameters underlying the fundamental beta. Table 3.2 presents summary statistics of the posterior distribution of the $δ_0$ and $δ_1$ parameters in equation (3.6). The constant term $δ_0$, which can be interpreted as the average fundamental beta because all conditioning variables are cross-sectionally demeaned, has a posterior mean of 1.01. The results show that all three firm characteristics are important determinants of beta.

\(^7\) Related to this, Engle and Rangel (2008) model low-frequency patterns in market volatility as a function of macroeconomic and financial variables and Hoberg and Welch (2007) compute long- and short-run betas based on different windows of past returns.
3.5 Empirical Results

Figure 3.1: Cross-Sectional Distribution of Phi

This figure shows the cross-sectional distribution of the time series average of the parameter $\phi_i$, which measures the proportion of beta explained by past realized beta,

$$\beta_{it} = \phi_{it} b_{it} + (1 - \phi_{it}) \beta^*_it,$$

where $b_{it}$ is the realized beta of firm $i$, $\beta^*_it$ is the fundamental beta, and where $\phi_{it}$ is given by

$$\phi_{it} = \phi_0 + \phi_1 V_{Mt},$$

where $V_{Mt}$ is the realized market variance. We first calculate $\phi_{it}$ based on equation (3.3) for each draw of the Gibbs sampler. We then calculate for each firm the time series average $\phi_i$ and its posterior mean. This figure shows the cross-sectional distribution of these posterior means.
Figure 3.2: Evolution of Phi through Time

This figure plots the evolution through time of the cross-sectional average of $\phi_{it}$. We first calculate at each iteration of the Gibbs sampler $\phi_{it}$ based on equation (3.3). We then compute its posterior mean and the cross-sectional average of these posterior means in each month from July 1964 through December 2006. Shaded areas indicate NBER recession periods.

![Cross-Sectional Mean of Phi](image.png)
3.5 Empirical Results

determinants of fundamental betas. Some of the interaction terms between the firm characteristics and macroeconomic variables also capture important variation in market betas, particularly those involving the default spread and one-month T-Bill rate.

Table 3.2: Posterior Distribution of Fundamental Beta Parameters

This table reports the Bayesian posterior distribution of the determinants of the fundamental beta, which is parameterized as a linear function of firm characteristics and business cycle variables,

$$\beta^*_t = \delta_0 + \delta_1[Z_{it} \otimes BC_t]$$

where $Z_{it}$ is a vector that contains $L$ firm characteristics and $BC_t$ is a vector that contains a constant and $M$ business cycle variables. MV is the log of firm size, BM is the log of the book-to-market ratio, and MOM is the cumulative return over the twelve months prior to the current month. These firm characteristics are expressed as deviations from their cross-sectional mean in every period. DEF is the default spread, DY is the dividend yield, TBILL is the one-month Treasury bill rate, and TERM is the term spread. The table presents the mean, median, standard deviation and 5th, 25th, 75th, and 95th percentile values of the posterior distribution of the delta parameters, based on 5,000 iterations of the Gibbs sampler and a burn-in period of 1,000 iterations. All $\delta_1$ parameters are multiplied by 100.

<table>
<thead>
<tr>
<th>Constant ($\delta_0$)</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>5th</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\delta_0$)</td>
<td>1.01</td>
<td>0.01</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td>MV</td>
<td>-3.62</td>
<td>0.68</td>
<td>-4.77</td>
<td>-4.05</td>
<td>-3.66</td>
<td>-3.20</td>
<td>-2.47</td>
</tr>
<tr>
<td>BM</td>
<td>-5.30</td>
<td>1.32</td>
<td>-7.38</td>
<td>-6.18</td>
<td>-5.44</td>
<td>-4.34</td>
<td>-3.10</td>
</tr>
<tr>
<td>MOM</td>
<td>-10.02</td>
<td>2.99</td>
<td>-14.97</td>
<td>-12.05</td>
<td>-9.99</td>
<td>-8.18</td>
<td>-5.28</td>
</tr>
<tr>
<td>MV*TBILL</td>
<td>0.65</td>
<td>0.29</td>
<td>0.16</td>
<td>0.45</td>
<td>0.65</td>
<td>0.84</td>
<td>1.14</td>
</tr>
<tr>
<td>MV*TERM</td>
<td>1.57</td>
<td>0.28</td>
<td>1.08</td>
<td>1.40</td>
<td>1.58</td>
<td>1.74</td>
<td>2.03</td>
</tr>
<tr>
<td>MV*DEF</td>
<td>0.55</td>
<td>0.52</td>
<td>-0.33</td>
<td>0.20</td>
<td>0.35</td>
<td>0.93</td>
<td>1.38</td>
</tr>
<tr>
<td>MV*DY</td>
<td>-1.41</td>
<td>0.08</td>
<td>-1.55</td>
<td>-1.47</td>
<td>-1.40</td>
<td>-1.35</td>
<td>-1.27</td>
</tr>
<tr>
<td>BM*TBILL</td>
<td>-3.11</td>
<td>0.58</td>
<td>-3.98</td>
<td>-3.52</td>
<td>-3.09</td>
<td>-2.74</td>
<td>-2.20</td>
</tr>
<tr>
<td>BM*TERM</td>
<td>0.35</td>
<td>0.73</td>
<td>-0.84</td>
<td>-0.13</td>
<td>0.39</td>
<td>0.83</td>
<td>1.49</td>
</tr>
<tr>
<td>BM*DEF</td>
<td>13.15</td>
<td>0.95</td>
<td>11.62</td>
<td>12.53</td>
<td>13.18</td>
<td>13.74</td>
<td>14.64</td>
</tr>
<tr>
<td>BM*DY</td>
<td>-0.05</td>
<td>0.14</td>
<td>-0.30</td>
<td>-0.14</td>
<td>-0.04</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>MOM*TBILL</td>
<td>3.05</td>
<td>1.02</td>
<td>1.43</td>
<td>2.29</td>
<td>3.06</td>
<td>3.81</td>
<td>4.67</td>
</tr>
<tr>
<td>MOM*TERM</td>
<td>-1.11</td>
<td>0.89</td>
<td>-2.55</td>
<td>-1.67</td>
<td>-1.08</td>
<td>-0.54</td>
<td>0.34</td>
</tr>
<tr>
<td>MOM*DEF</td>
<td>-28.91</td>
<td>1.70</td>
<td>-31.65</td>
<td>-30.02</td>
<td>-29.00</td>
<td>-27.76</td>
<td>-26.02</td>
</tr>
<tr>
<td>MOM*DY</td>
<td>8.18</td>
<td>0.23</td>
<td>7.81</td>
<td>8.03</td>
<td>8.18</td>
<td>8.32</td>
<td>8.56</td>
</tr>
</tbody>
</table>

As explained in section 3.2, we use the MIDAS approach of Ghysels, Santa-Clara, and Valkanov (2005) to estimate realized betas based on daily return data. This approach incorporates a flexible weighting function that makes it possible to choose the optimal weights given to past data in the estimation. The optimal window strikes a balance between giving equal weight to observations to obtain more precise beta estimates and giving more weight to recent data to obtain betas
that are timelier and therefore more relevant. As shown in equation (3.5), we use a beta weighting function whose shape is determined by two parameters. We set $\kappa_1$ equal to 1 and estimate $\kappa_2$. We find that in our realized beta specification the posterior mean of $\kappa_2$ is equal to 1.16. Figure 3.3 compares the optimal weighting scheme implied by the posterior mean of $\kappa_2$ to the equal weighting scheme used by rolling window estimators. The plot shows that in the optimal scheme the most recent 150 days receive more weight than in the equal weighting scheme because these are most informative for estimating realized betas.

We now turn to the mixed betas generated by our model. First, we calculate $\beta_{it}$ based on equation (3.2) at each iteration of the Gibbs sampler. Subsequently, we compute for every firm the time series average of the mixed beta and its posterior mean. Figure 3.4 shows the cross-sectional distribution of these posterior means of $\beta_i$. As expected, the distribution is centered around one and has a standard deviation of 0.34. A 95% confidence interval for beta ranges from 0.46 to 1.60, which implies that firms differ substantially in their sensitivity to broad market movements.

In Table 3.3 we report summary statistics of the posterior means of all three beta specifications. Because for each $t$ the cross-sectional average of $\bar{\beta}_i$ is close to one, the more interesting aspect is the dispersion in betas, both over time and in the cross-section. The left panel in Table 3.3 reports properties of the cross-section of betas and the right panel shows time series characteristics of beta. The diagonal elements in these panels show that on average, realized betas display the largest spread, both over time and across firms, while fundamental betas show the least variation. This is consistent with the notion that realized betas capture high-frequency movements in beta and fundamental betas pick up long-run beta fluctuations. Another explanation is that measurement error in the realized beta estimates leads to spurious dispersion or that in addition to firm size, book-to-market, and momentum, other firm characteristics drive variation in beta. The time series and cross-sectional behavior of the mixed betas is a combination of the dynamics of the realized and fundamental betas. Thus, it combines the benefits of both specifications, responding fast to changes in market conditions without producing excessive variation in beta. The off-diagonal elements in Table 3.3 are the correlations between the betas generated by the three specifications. Fundamental and realized betas are strongly correlated, both over time and across stocks. Correlation is far from perfect though, as a regression of one on the other has an $R^2$ of only 0.68. This illustrates that realized and fundamental betas exhibit different cross-sectional characteristics and time series dynamics. Hence, a combination of these two specifications captures different aspects of market beta dynamics.

### 3.5.2 Beta Estimation

In this section we compare the precision of beta estimates from the hierarchical Bayesian panel data model to that of estimates from a separate Bayesian time series regression for every firm. We study the estimation efficiency of the two methods for the fundamental beta specification, where $\phi_{it}$ is fixed at zero. Since
3.5 Empirical Results

Figure 3.3: Optimal versus Equal Weighting Scheme for Estimating Realized Beta

This figure compares the equal weights in the traditional rolling window estimator of realized betas to the weights implied by the MIDAS weighting function in equation (3.5) for the realized beta estimator in equation (3.4). We set the maximum window length equal to 250 trading days.
Figure 3.4: Cross-Sectional Distribution of Firm Betas

This figure shows the cross-sectional distribution of average firm betas. We first calculate at each iteration of the Gibbs sampler the beta for firm $i$ at time $t$ based on the model in equation (3.7). Subsequently, we compute the time series averages of these conditional betas. We then calculate for each firm the posterior mean of its time series average beta. This figure shows the cross-sectional distribution of these posterior means.
3.5 Empirical Results

Table 3.3: Beta Summary Statistics

This table reports summary statistics on the dispersion of betas and the correlation between mixed, fundamental, and realized betas. The left panel reports the properties of the cross-section of betas based on the time series average of the cross-sectional covariances

\[ S_{\text{cross},t} = \frac{1}{N_t} \sum_i \left( \bar{\beta}_{it}^{(j)} - \bar{\beta}_{it}^{(k)} \right) \left( \bar{\beta}_{it}^{(k)} - \bar{\beta}_{it}^{(s)} \right), \]

where the indices \( j \) and \( k \) refer to the model (Mixed, Realized, Fundamental), betas are evaluated at their posterior means \( \bar{\beta}_{it} \), and where \( \bar{\beta}_t \) is the average beta at time \( t \). The right panel considers the cross-sectional average of the time series covariances

\[ S_{\text{time},i} = \frac{1}{T_i} \sum_t \left( \bar{\beta}_{it}^{(j)} - \bar{\beta}_{i}^{(j)} \right) \left( \bar{\beta}_{it}^{(k)} - \bar{\beta}_{i}^{(k)} \right), \]

where \( \bar{\beta}_i \) is the average beta of firm \( i \). The diagonal elements in both panels have been transformed into standard deviations. The off-diagonal elements of both the cross-sectional and time series covariance matrices have been rescaled to correlations (stalics).

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional (( S_{\text{cross}} ))</th>
<th>Time Series (( S_{\text{time}} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed beta</td>
<td>0.33</td>
<td>0.24</td>
</tr>
<tr>
<td>Realized beta</td>
<td>0.95 0.61</td>
<td>0.94 0.46</td>
</tr>
<tr>
<td>Fundamental beta</td>
<td>0.94 0.83 0.10</td>
<td>0.93 0.82 0.09</td>
</tr>
</tbody>
</table>

this specification requires the estimation of many parameters when a large number of conditioning variables is included, the efficiency gain from using the panel model can be substantial. We measure estimation precision by computing the standard deviation and the 5% and 95% percentile values of the posterior distribution of beta at each point in time.

Figure 3.5 plots the posterior mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM from August 1964 through December 2006. The upper graph is based on the estimation output of the panel data model and the lower graph is constructed using the output of a time series regression. The shaded areas in the plot indicate NBER recession periods. The plots show that the confidence interval for beta obtained from the panel regression is much narrower than the interval produced by the time series regression. Noisy estimates of the \( \delta_1 \) parameters, which measure the influence of the conditioning variables on fundamental betas, lead to wide intervals for beta in the time series model.

The large efficiency gain in the panel model is due to two reasons. First, the \( \delta \) parameters are pooled across stocks in the panel specification. The panel model therefore exploits the information in the cross-section of stocks to obtain more precise estimates. Second, because we specify hierarchical priors for the firm-specific parameters in the model, only the parameters of the common distribution where the parameters are assumed to be drawn from have to be estimated. The Bayes estimator of the firm-specific parameters in the panel model shrinks the
3 Efficient Estimation of Firm-Specific Betas

Figure 3.5: Confidence Interval for IBM Beta: Panel versus Time Series Regression

This figure plots the mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM in each month from August 1964 through December 2006. The fundamental beta is modeled as a linear function of firm characteristics and macroeconomic state variables. The upper graph is based on the estimation output of the hierarchical panel data model presented in section 3.2 of the chapter and the lower graph is constructed using the output of a time series regression. Shaded areas indicate NBER recession periods.
least squares estimator towards the cross-sectional mean. In contrast, in the time series regressions every parameter is estimated individually, which results in poor estimation precision when many parameters need to be estimated and the number of time series observations is small.

Because in our panel approach parameters can be estimated more precisely, it can include more conditioning variables than the traditional approach of estimating a time series regression for every firm used by Avramov and Chordia (2006a). While we include 15 conditioning variables to accurately model beta dynamics, they note that “attention must be restricted to a small number of such variables to ensure some precision in the estimation procedure”. To compare the relative efficiency of the panel and time series approaches when a more parsimonious specification for fundamental betas is used, we also estimate the panel model and time series regressions with a set of conditioning variables that is similar to that used by Avramov and Chordia (2006a). In particular, we choose firm size, book-to-market, and two interactions terms between these characteristics and the default spread as instruments.

The confidence intervals for the fundamental beta of IBM based on this reduced set of conditioning variables are displayed in figure 3.6. As expected, the intervals for beta generated by the panel model and the time series regressions have both narrowed compared to those based on the complete set of conditioning variables. However, the plots show that even when less parameters need to be estimated the panel approach leads to more precise estimates of firm-specific betas than the time series approach.

Because IBM is present in our data set during the entire sample period, many observations are available for beta estimation (509 months). As explained before, we expect the efficiency gain from the hierarchical Bayesian panel data approach to be even larger for firms with a short return history. To summarize the estimation precision for the betas of all firms, we compute the cross-sectional average of the posterior standard deviations of all betas in every month. Figure 3.7 plots these standard deviations for the panel model and time series regressions. Clearly, the posterior standard deviation of betas estimated using the time series regressions is larger than the standard deviation of betas estimated using the panel regression.

3.5.3 Portfolio Heterogeneity

The previous section has shown that firm-specific betas are noisy when estimated using time series regressions. To reduce the measurement error in betas, Fama and MacBeth (1973) propose to aggregate stocks into portfolios and run a time series regression for every portfolio to obtain the portfolio’s beta. Fama and French (1992) follow this suggestion and assign each stock the beta of the portfolio it belongs to. As pointed out by Ferson and Harvey (1999), such an approach is often used in studies of initial public offerings (IPOs), when no return data is available to estimate beta. They note that this approach only works when the characteristics used for portfolio formation are good proxies for risk, because an important assumption underlying the portfolio approach is that the stocks in a
Figure 3.6: Confidence Interval for Reduced Fundamental Beta of IBM

This figure plots the mean and 5% and 95% percentile values of the posterior distribution of the fundamental beta of IBM in each month from August 1964 through December 2006. The fundamental beta is modeled as a linear function of a reduced set of firm characteristics and macroeconomic state variables. The upper graph is based on the estimation output of the hierarchical panel data model presented in section 3.2 of the chapter and the lower graph is constructed using the output of a time series regression. Shaded areas indicate NBER recession periods.
Figure 3.7: **Average Posterior Standard Deviation of Fundamental Betas**

This figure plots the cross-sectional average of the posterior standard deviations of the fundamental betas of all firms in the sample from August 1964 through December 2006. In the upper graph the fundamental beta is modeled as a linear function of firm characteristics and macroeconomic state variables and in the lower graph fundamental betas depend on a reduced set of conditioning variables. Posterior standard deviations are based on the estimation output of the hierarchical panel data model and the output of time series regressions estimated for every firm. Shaded areas indicate NBER recession periods.
particular portfolio share the same risk characteristics. In case of the widely used 25 portfolios sorted on firm size and book-to-market, it is assumed that firms are homogeneous in their exposure to risk after controlling for size and B/M. When the stocks in a given portfolio have different exposures to other determinants of risk, this method can lead to serious errors. In this section we therefore examine whether firms that are grouped together in a portfolio have similar risk characteristics.

We construct the 25 size-B/M portfolios following the procedure of Fama and French (1993).

Subsequently, we calculate for every portfolio \( j \) in every month \( t \) the cross-sectional average and standard deviation of the excess returns and posterior means of the alphas, betas, and phis of the stocks in that portfolio. The left part of Table 3.4 reports for each portfolio the time series means of these cross-sectional averages. Consistent with prior studies (e.g., Fama and French (1996)), the small-growth portfolio has the lowest average return and a large, negative pricing error. In general, the average portfolio returns display a strong value premium but weak size effect. Importantly, sorting on firm size and B/M does not produce a wide spread in average market betas across portfolios, as most portfolio betas are close to one.\(^8\)

The table further shows that the phi parameters of large cap portfolios are higher than those of small cap portfolios. This implies that realized betas are the most important determinants of the mixed betas of large firms while fundamental betas have a stronger effect on the mixed betas of small firms.

Table 3.4 also shows the dispersion of the risk and return characteristics across stocks in each portfolio. For all characteristics we observe strong heterogeneity within portfolios. In some portfolios the cross-sectional standard deviation of firm-specific alphas is more than 1%. Especially firms that are grouped together in small cap portfolios have significantly different pricing errors.

The cross-sectional variation in betas of firms in each portfolio is around 0.30, which implies that the assumption that stocks in the same portfolio have similar risk characteristics is violated. Table 3.4 also reports substantial heterogeneity in phi within portfolios. This means that for some firms in a given portfolio mixed betas are mainly driven by realized betas whereas for others fundamental betas are more important.

### 3.6 Applications of Firm-Specific Betas

This section discusses two important applications of the firm-level betas generated by our Bayesian panel data model. In section 3.6.1 we compare the explanatory power of different beta specifications and estimation methods for the cross-section of individual stock returns. Section 3.6.2 uses the beta forecasts to estimate the

---

\(^8\) However, unreported results show that value and growth portfolios exhibit very different risk dynamics. Confirming the results of Ang and Chen (2007) and Franzoni (2007), we find that the beta of value firms shows a declining trend and is lower than the beta of growth stocks since the 1980s.
Table 3.4: 

Heterogeneity within 25 Size-B/M Portfolios

This table presents the time series average of characteristics of 25 size-B/M sorted portfolios and the cross-sectional spread in these characteristics. The portfolios are constructed annually by sorting stocks independently into size and B/M quintiles at the end of June. The 25 portfolios are then formed as the intersections of these five size and B/M quintiles. Subsequently, we calculate for every portfolio \( j \) at every time \( t \) the equally weighted cross-sectional average of the posterior means of the firm-specific alphas, betas, and phis, and of the excess returns of the stocks in the portfolio. We also calculate for every portfolio \( j \) at every time \( t \) the cross-sectional standard deviation of the posterior means of the firm-specific alphas, betas, and phis, and of the excess returns of the stocks in the portfolio. The table shows for every portfolio the time series means of these cross-sectional averages and standard deviations.

<table>
<thead>
<tr>
<th>Size quintiles</th>
<th>Book-to-Market equity (B/M) quintiles</th>
<th>Average Return</th>
<th>Return Variation</th>
<th>Average Alpha</th>
<th>Alpha Variation</th>
<th>Average Beta</th>
<th>Beta Variation</th>
<th>Average Phi</th>
<th>Phi Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
</tr>
<tr>
<td>Small</td>
<td>0.18 0.54 0.67 0.97 0.93</td>
<td>15.56 13.84 13.11 12.50 13.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.72 0.83 0.80 0.83 0.85</td>
<td>12.98 11.16 10.40 10.13 11.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.67 0.69 0.76 0.95 0.87</td>
<td>11.03 9.83 9.04 8.89 9.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.65 0.62 0.72 0.84 0.94</td>
<td>9.26 8.26 7.83 7.51 8.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>0.48 0.46 0.59 0.72 0.84</td>
<td>7.27 6.93 6.57 6.35 6.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thesis_Rik V1.pdf

Thesis_Rik V2.pdf
covariance matrix of stock returns, which we then use to construct minimum variance portfolios.

3.6.1 Cross-Sectional Tests of the Conditional CAPM

The previous section has shown that aggregating individual stocks into portfolios leads to a substantial loss of information and shrinks the cross-sectional variation in betas. Ang, Liu, and Schwarz (2008) demonstrate that this loss of information can lead to large efficiency losses in cross-sectional tests of asset pricing models. In particular, they show that while creating portfolios reduces estimation error in betas, standard errors of risk premia estimates are higher due to the smaller spread in betas. Consequently, using individual stocks instead of portfolios as base assets allows for more powerful tests of asset pricing models.

Another important reason for using individual stocks in cross-sectional tests of asset pricing models is given by Lewellen, Nagel, and Shanken (2008). They show analytically that due to the strong factor structure in the 25 size-B/M sorted portfolios often used as test assets in asset pricing studies, traditional cross-sectional tests have low power to reject a model. In particular, when theoretical restrictions on cross-sectional slopes are ignored, any factor that is only weakly correlated with the true factors can generate high cross-sectional $R^2$'s and small pricing errors. Lewellen and Nagel (2006) show that the empirical support for several recently proposed asset pricing models weakens considerably when this issue is taken into account. Because individual stock returns do not have a strong factor structure, they are not affected by this problem.

In their analysis, Ang, Liu, and Schwarz (2008) assume constant stock betas, which they estimate by running time series regressions. We extend their work in two directions. First, we improve the specification of betas by allowing for time variation. Second, we use a formal panel data approach to increase the precision of firm-specific beta estimates. We do not claim that the CAPM is the “best” asset pricing model. Our objective is to show the effect of better beta specification and estimation on the pricing ability of the CAPM.

We first estimate betas for all stocks in our sample and for the 25 size-B/M portfolios. We consider four beta specifications (mixed, fundamental, realized, and static) and estimate the models using hierarchical Bayesian panel regressions. These betas are then used as independent variables in second stage monthly cross-sectional regressions of excess returns on betas,

$$r_{it} = \lambda_0t + \lambda_1t \beta_{it-1} + \lambda_2t x_{it-1} + \nu_{it}, \quad (3.10)$$

where $\lambda_0t$ is the intercept, $\lambda_1t$ the risk premium, and where $x_{it-1}$ is a vector of control variables. We run the cross-sectional regressions for every draw of the Gibbs sampler and calculate the time series average of the cross-sectional coefficients. We then calculate the posterior mean and variance of the Fama-MacBeth estimators. In appendix B we demonstrate that this procedure accounts for measurement error in beta by using the entire posterior distribution of the $\beta_{it}$ in the estimation.
Columns 1-3 in Table 3.5 report the Fama-MacBeth coefficient estimates when individual stocks are the test assets and no control variables are included in the regression \((x_{it-1} = 0)\). We find that for the mixed beta specification the intercept is close to zero and insignificant while the risk premium estimate is significantly positive. The \(\lambda_1\) estimate is 0.56%, which is close to the average monthly excess market return during the sample period (0.47%). This implies that the conditional CAPM with mixed betas satisfies the theoretical restriction emphasized by Lewellen and Nagel (2006) that the risk premium should equal the expected excess factor return. For the other three beta specifications the intercepts are significantly different from zero. The risk premium estimates in the realized beta and fundamental beta models are positive and significant but deviate more from the average market return than the premium estimated in the mixed beta model. In terms of explanatory power the mixed beta specification also outperforms the competing approaches to modeling beta. The static CAPM performs worst, because in this model the cross-sectional variation in market betas does not respond to business cycle variations.

Columns 4-6 in Table 3.5 show that when portfolios are used as test assets, all beta specifications generate economically large intercepts. Nevertheless, the mixed beta specification again has the highest explanatory power and the static CAPM does worst. We stress that the \(R^2\) should only be compared across beta specifications and should not be used to compare individual stocks and portfolios as test assets, because the dependent variables in the cross-sectional regressions are different. Table 3.5 further shows that the standard errors of the parameter estimates are much larger when portfolios are used as test assets than when individual stocks are used, which confirms that sorting stocks into portfolios can lead to large efficiency losses because it reduces the dispersion of betas. In fact, standard errors from using portfolios are more than twice as large as those from using individual stocks.

The last three columns in Table 3.5 report estimation results for individual stocks when control variables are added to the cross-sectional regressions. In particular, the vector \(x_{it-1}\) contains the firm characteristics size, book-to-market, and momentum. Fama and French (1992) find that the cross-sectional relation between market beta and average return is flat when tests control for size. We find that while adding these firm characteristics leads to an increase in explanatory power, the risk premium estimate for the mixed beta specification remains significantly positive. Thus, when individual stocks are used as test assets and betas are well-specified and precisely estimated, the positive relation between beta and return no longer disappears when controlling for firm characteristics.

### 3.6.2 Beta Forecasts and Minimum Variance Portfolios

An important application of betas is to estimate the covariance matrix of returns, which is used to construct mean-variance efficient portfolios. Traditional implementations of the portfolio theory developed by Markowitz (1952) use sample moments. When the number of assets is large, however, it is difficult to precisely...
Table 3.5: Cross-Sectional Asset Pricing Tests

This table reports Fama-MacBeth coefficient estimates for 5,017 NYSE-AMEX stocks (columns 1-3 and 7-9) and for 25 size-B/M portfolios (columns 4-6). The sample period is August 1964 through December 2006 (509 months). In the first stage betas are estimated using the Bayesian panel data approach outlined in section 3.2. These betas are then used as independent variables in second stage cross-sectional regressions of returns on betas, 
\[ r_{it} = \lambda_0 t + \lambda_1 \beta_{it} - 1 + \lambda_2 x_{it} - 1 + \upsilon_{it}. \]

In columns 1-6 no control variables are included (\( x_{it} - 1 = 0 \)). In columns 7-9 the vector \( x_{it} - 1 \) includes the firm characteristics size, book-to-market, and momentum. We use the entire posterior distribution of beta in the estimation to account for measurement error in beta. The table shows the estimated intercept (\( \lambda_0 \)) and risk premium (\( \lambda_1 \)) for four different beta specifications. Standard errors are in parentheses and corresponding t-stats are reported in brackets.

<table>
<thead>
<tr>
<th>Panel A: Mixed Beta</th>
<th>Panel B: Realized Beta</th>
<th>Panel C: Fundamental Beta</th>
<th>Panel D: Total Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stocks - No controls</td>
<td>Stocks - With controls</td>
<td>Portfolios - No controls</td>
<td>Portfolios - With controls</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>( \lambda_1 )</td>
<td>Adj. R^2 (%)</td>
<td>( \lambda_0 )</td>
</tr>
<tr>
<td>0.04</td>
<td>0.56</td>
<td>4.21</td>
<td>0.61</td>
</tr>
<tr>
<td>(0.17)</td>
<td>(0.26)</td>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>[-0.24]</td>
<td>[2.15]</td>
<td></td>
<td>[1.45]</td>
</tr>
<tr>
<td>Panel A: Mixed Beta</td>
<td>Panel B: Realized Beta</td>
<td>Panel C: Fundamental Beta</td>
<td>Panel D: Total Beta</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>( \lambda_1 )</td>
<td>Adj. R^2 (%)</td>
<td>( \lambda_0 )</td>
</tr>
<tr>
<td>0.35</td>
<td>0.35</td>
<td>3.47</td>
<td>0.61</td>
</tr>
<tr>
<td>(0.13)</td>
<td>(0.16)</td>
<td></td>
<td>(0.29)</td>
</tr>
<tr>
<td>[2.73]</td>
<td>[2.24]</td>
<td></td>
<td>[2.07]</td>
</tr>
<tr>
<td>Panel A: Mixed Beta</td>
<td>Panel B: Realized Beta</td>
<td>Panel C: Fundamental Beta</td>
<td>Panel D: Total Beta</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>( \lambda_1 )</td>
<td>Adj. R^2 (%)</td>
<td>( \lambda_0 )</td>
</tr>
<tr>
<td>0.64</td>
<td>0.67</td>
<td>2.05</td>
<td>1.14</td>
</tr>
<tr>
<td>(0.30)</td>
<td>(0.34)</td>
<td></td>
<td>(0.76)</td>
</tr>
<tr>
<td>[2.13]</td>
<td>[1.97]</td>
<td></td>
<td>[1.50]</td>
</tr>
<tr>
<td>Panel A: Mixed Beta</td>
<td>Panel B: Realized Beta</td>
<td>Panel C: Fundamental Beta</td>
<td>Panel D: Total Beta</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>( \lambda_1 )</td>
<td>Adj. R^2 (%)</td>
<td>( \lambda_0 )</td>
</tr>
<tr>
<td>0.63</td>
<td>0.09</td>
<td>0.65</td>
<td>1.22</td>
</tr>
<tr>
<td>(0.28)</td>
<td>(0.20)</td>
<td></td>
<td>(0.47)</td>
</tr>
<tr>
<td>[2.25]</td>
<td>[0.45]</td>
<td></td>
<td>[2.60]</td>
</tr>
</tbody>
</table>
estimate the expected returns and covariances. As a result, asset weights are often extreme and portfolios behave poorly out-of-sample. Many strategies have been proposed to improve the out-of-sample performance of mean-variance portfolios, including shrinkage estimators, imposing short-selling constraints, using asset pricing models to estimate expected returns, and imposing a factor structure on the covariance matrix.\footnote{See, e.g., Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003), and Ledoit and Wolf (2003).}

In a recent study, DeMiguel, Garlappi, and Uppal (2009) compare the out-of-sample performance of these approaches to the $1/N$ rule that gives equal weight to all available assets. They conclude that none of the more sophisticated methods consistently outperforms the naive $1/N$ benchmark in terms of Sharpe ratio or certainty equivalent return. Of the alternative models considered, the minimum variance portfolio with short-selling constraints proposed by Jagannathan and Ma (2003) has the highest Sharpe ratio but is not superior to the $1/N$ strategy. The global minimum variance portfolio does well, because it is the only efficient portfolio that does not require estimates of expected returns, which contain large estimation errors. Chan, Karceski, and Lakonishok (1999) compare the out-of-sample performance of minimum variance portfolios based on forecasts of future covariances produced by factor models. They find that there is one major factor, the market, that dominates all other factors. Hence, the one-factor model is adequate for forming the global minimum variance portfolio.

Motivated by these findings, we use the mixed beta forecasts produced by our Bayesian panel data model to forecast the covariance matrix of stock returns and construct the global minimum variance portfolio. We expect our method to outperform competing approaches because it delivers more efficient estimates of firm-specific betas and because it allows for time variation in beta. DeMiguel, Garlappi, and Uppal (2009) admit that their assumption of constant risk is a limitation, but argue that models that allow for time-varying moments are likely to perform poorly out-of-sample because many parameters need to be estimated. However, one of the key advantages of our method is that it can estimate many parameters with high precision. We compare the out-of-sample performance of our approach to that of the traditional sample covariance matrix, the static one-factor structure considered by Chan, Karceski, and Lakonishok (1999), and the $1/N$ rule advocated by DeMiguel, Garlappi, and Uppal (2009).\footnote{Brandt, Santa-Clara, and Valkanov (2009) propose a new approach to portfolio optimization, in which portfolio weights are modeled as a function of firm characteristics to exploit cross-sectional patterns in stock returns. They show that this parametric portfolio policy performs well for expected return maximizing portfolios but note that it works less well when the objective is to construct risk minimizing portfolios.} Engle and Colacito (2006) stress the importance of isolating the effect of covariance information from expected returns when the objective is to evaluate different covariance estimators. Because expected returns do not enter the optimization when constructing minimum variance portfolios, differences in the portfolio weights only reflect the effect of different covariance forecasts.
The first estimator of the covariance matrix is the sample covariance matrix,
\[ S_t^S = \frac{1}{T-1} \sum_{i=1}^{T} (R_t - \bar{R})(R_t - \bar{R}), \]  
\[ (3.11) \]
where \( R_t \) is the vector of monthly stock returns and \( \bar{R} \) contains the sample mean returns. Our second covariance estimator is based on the one-factor model. In the first step, we estimate mixed betas using our panel approach and static betas using time series regressions. We use these betas to estimate the covariance matrix in each month according to the one-factor model,
\[ S_t^F = s_{Mt}^2 B_t B_t' + D, \]  
\[ (3.12) \]
where \( B_t \) is the \( N_t \times 1 \) vector of betas, \( s_{Mt}^2 \) is the sample variance of the market premium, and \( D \) is a diagonal matrix that contains the variances of the residuals.

Subsequently, we use the various estimates of the covariance matrix to construct the minimum variance portfolio, by choosing the portfolio weights that solve the following problem:
\[ \min w_t' S_t w_t, \]  
\[ \text{s.t. } \sum_i w_{it} = 1. \]  
\[ (3.13) \]
\[ (3.14) \]

The constraint implies that the portfolio is fully invested. Following Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2003), we also consider an extension in which we add a short-selling constraint,
\[ w_{it} \geq 0, \quad i = 1, 2, \ldots N. \]  
\[ (3.15) \]

We form the minimum variance portfolio at the end of each month, based on the forecast of the covariance matrix for the next month.

The first portfolio is formed using the first half of the sample period to forecast the covariance matrix of returns according to the methods explained above. Because the sample covariance matrix cannot be positive definite unless the number of return observations per stock is larger than the number of stocks, we only apply this method to a subset of the investment universe. We record the performance of the minimum variance portfolio in the next month and rebalance the portfolio using the new forecast of the covariance matrix. This method produces a time series of monthly returns for global minimum variance portfolios constructed using different covariance estimators. As a benchmark we also form an equally weighted portfolio at the end of each month. Since the objective is to minimize the portfolio variance, we evaluate the performance of the different methods by calculating the realized volatility of the portfolio returns.

Table 3.6 reports annualized risk and return characteristics of the minimum variance portfolios constructed using various forecasts of the covariance matrix. Panel A shows the out-of-sample performance when all stocks in the sample are
3.6 Applications of Firm-Specific Betas

used in the optimization and short-selling is allowed. The mixed beta specification estimated using the panel data approach outperforms all other methods and produces a portfolio with an annualized standard deviation of 8.12%. The 1/N strategy leads to a standard deviation that is almost twice as large (15.40%). The static beta model ranks second and generates an out-of-sample standard deviation of 8.50%. Although our objective is to minimize portfolio variance, we find that the mixed beta approach also leads to the best risk-return tradeoff, as it produces the highest Sharpe ratio of all four approaches, equal to 0.69. Since the minimum variance portfolio constructed using the mixed beta approach does not involve taking short positions, it can also be easily implemented in practice. No short positions are taken because the mixed beta forecasts for all stocks are positive.

In panel B we report the performance for portfolios constructed from a random sample of 250 stocks, which is the same number of stocks considered by Chan, Karceski, and Lakonishok (1999). For this subset of stocks the 1/N rule leads to the highest out-of-sample standard deviation (15.93%), followed by the sample covariance matrix, which generates a standard deviation of 15.19% and involves taking large short positions. The mixed beta panel approach again beats all other methods and produces a portfolio with a standard deviation of 8.41%. Thus, the performance of this approach when applied to the smaller sample of stocks is similar to that when all stocks are used in the optimization. An important reason for the relatively poor performance of the naive 1/N strategy is that we allocate wealth across individual stocks. In contrast, DeMiguel, Garlappi, and Uppal (2009) apply this policy to allocate wealth across portfolios of stocks. They point out that the loss from naive as opposed to optimal diversification is much larger when allocating wealth across individual assets, because individual stocks have higher idiosyncratic volatility than portfolios.

Jagannathan and Ma (2003) document that the out-of-sample performance of the sample covariance matrix can be improved by imposing no-short-sale constraints, because these reduce sampling error. Panel C reports the risk and return characteristics of the portfolios generated by the four methods when the nonnegativity constraint is imposed on the weights. The random sample of 250 stocks used to form the portfolios is the same as that used in panel B. We find that the standard deviation of the minimum variance portfolio constructed using the sample covariance matrix is indeed lower when no-short-sale restrictions are in place. In fact, this method yields a lower standard deviation than the equally weighted (1/N) portfolio or the one-factor model with static betas. We also confirm the finding of Jagannathan and Ma (2003) that imposing no-short-sale constraints reduces the performance of the static factor model. Because the portfolio produced by the mixed beta approach in panel B does not take short positions, it is not affected by the nonnegativity constraint and still has the smallest out-of-sample standard deviation.

63
Table 3.6: Risk and Return Characteristics of Global Minimum Variance Portfolios

This table reports the out-of-sample performance of global minimum variance portfolios that are formed at the end of each month from December 1985 through December 2006 out of a universe of NYSE-AMEX stocks. The optimization procedure uses forecasts of the covariance matrix of returns produced by different models. Panel A reports the out-of-sample performance of these minimum variance portfolios when all stocks in the sample are used in the optimization and without any constraints imposed on the weights. Panel B reports results for this unconstrained optimization when a random sample of 250 stocks is used to construct the portfolios. Panel C reports results for this reduced investment universe when a nonnegativity constraint is imposed on the portfolio weights (no short-selling).

The mean return and standard deviation are those of excess returns. Mean return, standard deviation, and Sharpe ratio are annualized. Short interest is in percentages.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Sharpe Ratio</th>
<th>Short Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Unconstrained (all stocks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equally weighted (1/N)</td>
<td>8.57</td>
<td>15.40</td>
<td>0.56</td>
<td>0.00</td>
</tr>
<tr>
<td>Static beta (TS model)</td>
<td>5.24</td>
<td>8.50</td>
<td>0.61</td>
<td>-64.64</td>
</tr>
<tr>
<td>Mixed beta (Panel model)</td>
<td>5.60</td>
<td>8.12</td>
<td>0.69</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Panel B: Unconstrained (250 stocks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample covariance matrix</td>
<td>4.30</td>
<td>15.19</td>
<td>0.28</td>
<td>-144.79</td>
</tr>
<tr>
<td>Equally weighted (1/N)</td>
<td>9.46</td>
<td>15.93</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Static beta (TS model)</td>
<td>7.81</td>
<td>13.32</td>
<td>0.59</td>
<td>-60.35</td>
</tr>
<tr>
<td>Mixed beta (Panel model)</td>
<td>5.70</td>
<td>8.41</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Panel C: Nonnegativity Constrained (250 stocks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample covariance matrix</td>
<td>3.02</td>
<td>11.74</td>
<td>0.26</td>
<td>0.00</td>
</tr>
<tr>
<td>Equally weighted (1/N)</td>
<td>9.46</td>
<td>15.93</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Static beta (TS model)</td>
<td>7.55</td>
<td>15.60</td>
<td>0.48</td>
<td>0.00</td>
</tr>
<tr>
<td>Mixed beta (Panel model)</td>
<td>5.70</td>
<td>8.41</td>
<td>0.68</td>
<td>0.00</td>
</tr>
</tbody>
</table>
3.7 Conclusion

Many applications of modern finance theory require precise beta estimates for individual stocks.

However, as noted by Campbell, Lettau, Malkiel, and Xu (2001), “firm-specific betas are difficult to estimate and may well be unstable over time”. Academics and practitioners have taken two approaches to estimating firm-level betas. The first method sorts stocks into portfolios based on characteristics to reduce measurement error and assigns each firm the beta of the portfolio it belongs to. However, when stocks in the same portfolio have different exposures to other determinants of risk than the characteristics they are sorted on, this approach can lead to serious errors. The second method estimates a separate time series regression for every stock. Although this approach allows each firm to have a different risk exposure, the resulting beta estimates can be very noisy. The literature also uses different specifications to model time variation in betas. Many studies use a parametric approach in which variation in beta is modeled as a linear function of conditioning variables. An alternative, non-parametric approach to model risk dynamics is based on purely data-driven filters. However, both methods have important drawbacks and involve a trade-off between precision and timeliness of beta estimates.

In this chapter we therefore improve both the specification and estimation of firm-specific, time-varying betas. We combine the parametric and non-parametric approaches for modeling changes in betas. The precision of firm-level beta estimates is increased by setting up a Bayesian panel data model that exploits the information contained in the cross-section of stocks and imposes a common structure on parameters while still allowing for cross-sectional heterogeneity.

We find that modeling time-varying betas as a function of both conditioning variables and past return data is preferred over traditional beta specifications that are based on only one of these components. Because fundamental and realized betas exhibit different time series dynamics and cross-sectional characteristics, a combination of these specifications captures different aspects of beta. We show that the optimal mixture of these two betas varies across firms and over time.

We further demonstrate that our panel data approach yields more precise estimates of firm-level betas than the traditional approach of estimating betas by running a time series regression for every firm. Moreover, we document strong cross-sectional variation in betas of firms that are grouped together in portfolios sorted on size and book-to-market. Consequently, aggregating stocks into portfolios conceals important information contained in individual stock betas and reduces the cross-sectional variation in betas.

We demonstrate that the mixed betas generated by our panel data model lead to a sharp increase in the pricing ability of the conditional CAPM. The estimate of the market risk premium remains significantly positive when controlling for firm characteristics. The results support the finding of Ang, Liu, and Schwarz (2008) that the use of individual stocks as test assets instead of portfolios leads to more efficient estimates in cross-sectional tests of asset pricing models. We extend their work by showing that a better specification and more precise estimation of stock-
specific betas increases the explanatory power of the CAPM.

Accurate estimates of firm-specific betas are also important for portfolio optimization. Based on the mixed beta estimates produced by our panel model we forecast the covariance matrix of stock returns, which is then used to form minimum variance portfolios. The portfolio constructed using mixed betas from the Bayesian panel approach outperforms portfolios of other strategies, such as the traditional sample covariance matrix and the naive $1/N$ rule, in terms of out-of-sample standard deviation. The mixed beta specification is also superior to competing approaches when short-selling constraints on portfolio weights are imposed.

Since our framework is flexible, it can be readily extended to include multiple risk factors, a different set of conditioning variables for fundamental betas, or another window length for estimating realized betas. In addition, while we have demonstrated the advantages of our approach for asset pricing and portfolio management, it also has important benefits in corporate finance applications. Specifically, because it quickly captures changes in beta and generates precise beta estimates even when little return data is available, our method is well suited for calculating risk-adjusted returns in studies of IPOs and M&As.
3.A Appendix I: Posterior Distributions

3.A.1 Joint Posterior Distribution

The joint posterior density is proportional to the product of the likelihood function and the prior distributions of all parameters \( \theta \): 

\[
p(\theta | y) \propto p(y | \theta) p(\theta).
\]

Defining \( \beta_{it} \) as in equation (3.2), stacking the time series observations for every firm \( i \) into vectors, and substituting the prior densities specified in section 3.3 produces the following joint posterior distribution\(^{11}\):

\[
p(\theta_i | y) = p(\alpha_i, \sigma^2 \alpha_i, \phi_i, \sigma^2 \phi_i, \delta_i, \kappa_2, \delta_0, \sigma^2 \delta_0, \delta_1, \Omega^{\delta_1 - 1}_i, \sigma^2 \epsilon_i | y)
\]

\[
\propto \prod_{i=1}^{N} \left( \sigma^2 \epsilon_i \right)^{-\frac{\tau}{2}} \exp \left[ -\frac{1}{2 \sigma^2 \epsilon_i} (r_i - \alpha_i - \kappa_2 \delta_i) (r_i - \alpha_i - \kappa_2 \delta_i) \right]
\]

\[
\times \prod_{i=1}^{N} \left( \sigma^2 \alpha_i \right)^{-\frac{\beta}{2}} \exp \left[ -\frac{\beta}{2 \sigma^2 \alpha_i} \right] \times \left( \sigma^{-2} \alpha_i \right)^{A_{\alpha}+1} \exp \left[ -\sigma^{-2} \alpha_i \right]
\]

\[
\times \prod_{i=1}^{N} \left( \sigma^2 \phi_i \right)^{-\frac{\phi_0}{2}} \exp \left[ -\left( \phi_0 - \mu \phi_0 \right)^2 \right] \times \left( \sigma^{-2} \phi_0 \right)^{A_{\phi}+1} \exp \left[ -\sigma^{-2} \phi_0 \right]
\]

\[
\times \left( \sigma^2 \phi_i \right)^{-\frac{\phi_0}{2}} \exp \left[ -\left( \phi_0 - \mu \phi_0 \right)^2 \right] \times \left( \sigma^{-2} \phi_0 \right)^{A_{\phi}+1} \exp \left[ -\sigma^{-2} \phi_0 \right]
\]

\[
\times \left( \sigma^2 \delta_i \right)^{-\frac{\delta_0}{2}} \exp \left[ -\left( \delta_0 - \mu \delta_0 \right)^2 \right] \times \left( \sigma^{-2} \delta_0 \right)^{A_{\delta}+1} \exp \left[ -\sigma^{-2} \delta_0 \right]
\]

\[
\times \prod_{i=1}^{N} \left( |\Omega^{\delta_1 - 1}_i| \right)^{\frac{1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \psi_i, S^{\delta_1}_i, \Omega^{\delta_1}_i \right) \right) \right] \prod_{i=1}^{N} \left( \sigma^2 \epsilon_i \right)^{A_{\epsilon}+1} \exp \left[ -\sigma^{-2} \epsilon_i \right].
\]

3.A.2 Conditional Posterior Distributions

In order to implement the Gibbs sampler we need to derive the full conditional posterior densities for each block of parameters.

\(^{11}\) We use the following parametrization of the inverse gamma distribution,

\[
p(y | A, B) = B^A \frac{1}{\Gamma(A)} \left( \frac{1}{y} \right)^{A+1} \exp \left( -\frac{B}{y} \right),
\]

where \( \Gamma(A) \) denotes the Gamma function, \( A \) is the shape parameter, and \( B \) is the scale parameter. For the Wishart distribution we use the parameterization,

\[
p(H | R, \nu) \propto \frac{|H|^{(\nu-k-1)/2}}{|R|^{\nu/2}} \exp \left[ -\frac{1}{2} \text{tr} \left( R^{-1} H \right) \right],
\]

where \( k \) denotes the dimension of the matrix \( H \), \( \nu \) is the degrees of freedom parameter, and \( R \) is the scale matrix.
The conditional densities can be derived from the joint posterior density by ignoring all terms that do not depend on the parameters of interest and then treating the parameters considered to be known as constants. We then obtain the conditional density for the parameters of interest by rearranging the remaining terms into the kernel of a known distribution.

We partition the parameter vector $\theta$ into the following blocks:

\[ \theta^{(1)}: \text{MIDAS weight parameters: } (\kappa_2) \]

\[ \theta^{(2)}: \text{Alpha parameters: } (\alpha_i) \]

\[ \theta^{(3)}: \text{Fundamental beta parameters: } (\delta_0, \delta_1) \]

\[ \theta^{(4)}: \text{Firm-specific mixed beta parameters: } (\phi_{0i}) \]

\[ \theta^{(5)}: \text{Pooled mixed beta parameter: } (\phi_1) \]

\[ \theta^{(6)}: \text{Variance and covariance parameters: } (\sigma^2_\alpha, \sigma^2_{\delta_0}, \Omega^{-1}_{\delta_1}, \sigma^2_{\phi_0}, \sigma^2_{\phi_1}, \sigma^2_\epsilon) \]

To generate samples from the conditional posterior of $\theta^{(1)}$ we use the Metropolis-Hastings algorithm. The conditional posteriors for all other blocks have convenient functional forms. Therefore, we use the Gibbs sampler to iteratively draw from the conditional densities of $\theta^{(2)}, \theta^{(3)}, \theta^{(4)}, \theta^{(5)}, \theta^{(6)}$.

To simplify notation we rewrite the model in matrix form as

\[
\begin{align*}
    \mathbf{r}_i &= \alpha_i \mathbf{T}_i + \phi_0 \mathbf{r}_M \mathbf{b}_i + \phi_1 \mathbf{r}_M \mathbf{V}_M \mathbf{b}_i + \mathbf{r}_M \mathbf{V}_M \mathbf{W}_i \delta - \phi_0 \mathbf{r}_M \mathbf{W}_i \delta - \phi_1 \mathbf{r}_M \mathbf{V}_M \mathbf{W}_i \delta + \epsilon_i,
\end{align*}
\]  

(3.16)

where $\mathbf{r}_i$ is an $T \times 1$ vector of excess returns, $\mathbf{r}_M$ an $T \times T$ diagonal matrix of excess market returns, $\mathbf{V}_M$ an $T \times T$ diagonal matrix of lagged market volatility, $\mathbf{b}_i$ an $T \times 1$ vector of realized betas, and $\epsilon_i$ an $T \times 1$ vector of idiosyncratic shocks. Since the $\delta_0$ and $\delta_1$ parameters in block $\theta^{(3)}$ have independent priors, we have simplified the notation further by rewriting $\delta_0 \mathbf{T}_i + ZBC_i \delta_1$ as $\mathbf{W}_i \delta$, where $\mathbf{W}_i$ is the $T \times (1 + L + LM)$ matrix of the constant term and conditioning variables. We combine the corresponding precisions $\sigma^{-2}_{\delta_0}$ and $\Omega^{-1}_{\delta_1}$ into the matrix $\Omega^{-1}_{\delta}$.

### 3.A.3 Metropolis-Hastings algorithm to draw $\kappa_2$

Since we implement a change of variable, $\kappa_2 = 1 + 25 \tilde{\kappa}_2$, we need to draw values for $\tilde{\kappa}_2$. Because the conditional posterior density for $\tilde{\kappa}_2$ does not take a standard form, we cannot use the Gibbs sampler. Instead, we employ the Metropolis-Hastings algorithm, which is a general accept-reject algorithm. In fact, Gelman, Carlin, Stern, and Rubin (2004) show that the Gibbs sampler is a special case of Metropolis-Hastings in which proposed parameter values are accepted with probability one. The M-H algorithm proceeds as follows.

First, a candidate value $\tilde{\kappa}_2^*$ is drawn from a proposal density $q(\tilde{\kappa}_2)$. We apply the Independence Chain M-H algorithm, in which the proposal density is independent across draws. We choose a Beta(1,3) proposal density, which has a mean of 0.25 and standard deviation equal to 0.19. Because the proposal density is not identical to the posterior density, the M-H algorithm does not accept all proposal draws. When a proposal is rejected the parameter value is set equal to the current
value. Draws are accepted according to the following probability

\[ \pi(\tilde{\kappa}_2^{(g-1)}, \tilde{\kappa}_2^*) = \min \left\{ 1, \frac{p(\tilde{\kappa}_2^*) q(\tilde{\kappa}_2^{(g-1)})}{p(\tilde{\kappa}_2^{(g-1)}) q(\tilde{\kappa}_2^*)} \right\}. \] (3.17)

This approach ensures that candidate draws with a high posterior density have a higher probability of being accepted than draws with a low posterior density. Repeating this procedure produces the required sequence of draws from the posterior distribution.

### 3.A.4 Conditional Posterior \( \alpha_i \)

Using Bayes’ theorem, we can write:

\[ p(\alpha_i | y) \propto p(y | \alpha_i) p(\alpha_i) \]

\[ \propto \exp \left[ -\frac{1}{2} Q^* \right], \]

where

\[ Q^* = (r_i - \alpha_i - r_M \beta_i)' \Omega_{\epsilon_i}^{-1} (r_i - \alpha_i - r_M \beta_i) + \frac{\alpha_i^2}{\sigma^2_{\alpha_i}} \]

\[ = (X_{\alpha_i} - \alpha_i \iota_T)' \Omega_{\epsilon_i}^{-1} (X_{\alpha_i} - \alpha_i \iota_T) + \frac{\alpha_i^2}{\sigma^2_{\alpha_i}} \]

\[ = Q_1^* + Q_2^*, \]

with

\[ Q_1^* = \frac{(\alpha_i - \bar{\alpha}_i)^2}{\sigma^2_{\alpha_i}} , \]

and

\[ Q_2^* = X_{\alpha_i}' \Omega_{\epsilon_i}^{-1} X_{\alpha_i} - \frac{\alpha_i^2}{\sigma^2_{\alpha_i}} . \]

and where

\[ X_{\alpha_i} = r_i - \phi_0 r_M b_i - \phi_1 r_M V_M b_i - r_M W_i \delta \]

\[ + \phi_0 r_M W_i \delta + \phi_1 r_M V_M W_i \delta. \]

We specify \( \theta^{(\alpha_i)} = \left( \sigma^2_{\alpha_i}, \phi_0, \phi_1, \sigma^2_{\alpha_i}, \alpha_2, \delta_0, \sigma^2_{\alpha_i}, \delta_1, \Omega_{\epsilon_i}^{-1}, \sigma^2_{\epsilon} \right) \). In the derivation of \( p(\alpha_i | \theta^{(\alpha_i)}, y) \), all parameters in \( Q_2^* \) are known, so we can treat \( Q_2^* \) as a constant. Thus, \( p(\alpha_i | \theta^{(\alpha_i)}, y) \) is proportional to \( \exp[-\frac{1}{2} Q_1^*] \), which is the kernel of a normal density. Therefore,

\[ \alpha_i | \theta^{(\alpha_i)}, y \sim N \left( \bar{\alpha}_i, \sigma^2_{\alpha_i} \right), \]

with

\[ \bar{\alpha}_i = \left[ \nu_T' \Omega_{\epsilon_i}^{-1} \nu_T + \frac{1}{\sigma^2_{\alpha_i}} \right]^{-1} \left[ \nu_T' \Omega_{\epsilon_i}^{-1} X_{\alpha_i} \right], \]

and

\[ \sigma^2_{\alpha_i} = \left[ \nu_T' \Omega_{\epsilon_i}^{-1} \nu_T + \frac{1}{\sigma^2_{\alpha_i}} \right]^{-1} . \]
3 Efficient Estimation of Firm-Specific Betas

3.A.5 Conditional Posterior \( \delta \)

\[
\delta | \theta, y \sim N(\bar{\delta}, \bar{\Omega}_\delta),
\]

with

\[
\bar{\delta} = \left[ \sum_{i=1}^{N} ((1 - \phi_{0i}) r_M W_i - \phi_1 r_M V_M W_i)' \Omega^{-1}_{\varepsilon_i} \right]^{-1} \\
\times \left[ \sum_{i=1}^{N} \left( (1 - \phi_{0i}) r_M W_i - \phi_1 r_M V_M W_i \right)' \right]^{-1} \\
\times \left[ \sum_{i=1}^{N} \Omega^{-1}_{\varepsilon_i} \left( (1 - \phi_{0i}) r_M W_i - \phi_1 r_M V_M W_i \right) \right],
\]

and

\[
\bar{\Omega}_\delta = \left[ \sum_{i=1}^{N} ((1 - \phi_{0i}) r_M W_i - \phi_1 r_M V_M W_i)' \Omega^{-1}_{\varepsilon_i} \right]^{-1} \\
\times \left[ \sum_{i=1}^{N} \left( (1 - \phi_{0i}) r_M W_i - \phi_1 r_M V_M W_i \right)' \right]^{-1},
\]

and where

\[
X_{\delta i} = r_i - \alpha_i T_i - \phi_0 r_M b_i - \phi_1 r_M V_M b_i.
\]

3.A.6 Conditional Posterior \( \phi_{0i} \)

\[
\phi_{0i} | \theta, y \sim N(\bar{\phi}_{0i}, \bar{\sigma}^2_{\phi_{0i}}),
\]

with

\[
\bar{\phi}_{0i} = \left[ (b_i - W_i \delta)' r_M \Omega^{-1}_{\varepsilon_i} r_M (b_i - W_i \delta) + \frac{1}{\bar{\sigma}^2_{\phi_{0i}}} \right]^{-1} \\
\times \left[ (b_i - W_i \delta)' r_M \Omega^{-1}_{\varepsilon_i} X_{\phi_{0i}} + \frac{\mu_{\phi_{0i}}}{\bar{\sigma}^2_{\phi_{0i}}} \right],
\]

and

\[
\bar{\sigma}^2_{\phi_{0i}} = \left[ (b_i - W_i \delta)' r_M \Omega^{-1}_{\varepsilon_i} r_M (b_i - W_i \delta) + \frac{1}{\bar{\sigma}^2_{\phi_{0i}}} \right]^{-1},
\]

and where

\[
X_{\phi_{0i}} = r_i - \alpha_i T_i - r_M W_i \delta - \phi_1 r_M V_M (b_i - W_i \delta).
\]
3.A.7 Conditional Posterior \( \phi_1 \)

\[
\phi_1 | \theta^{(1:5)}, y \sim N(\bar{\phi}_1, \sigma^2_{\phi_1}),
\]

with \( \bar{\phi}_1 = \left[ \sum_{i=1}^{N} (b_i - W_i \delta)^\prime r_M V_M \Omega^{-1}_{\epsilon_1} r_M V_M (b_i - W_i \delta) + \frac{1}{\sigma^2_{\phi_1}} \right]^{-1} \times \left[ \sum_{i=1}^{N} X'_{\phi_1} \Omega^{-1}_{\epsilon_1} r_M V_M (b_i - W_i \delta) \right],
\]

and \( \sigma^2_{\phi_1} = \left[ \sum_{i=1}^{N} (b_i - W_i \delta)^\prime r_M V_M \Omega^{-1}_{\epsilon_1} r_M V_M (b_i - W_i \delta) + \frac{1}{\sigma^2_{\phi_1}} \right]^{-1}, \)

and where \( X_{\phi_1} = r_i - \alpha_i x_i - \phi_0 r_M (b_i - W_i \delta) - r_M W_i \delta. \)

3.A.8 Conditional Posteriors \( \sigma^2_{\alpha}, \sigma^2_{\phi_0}, \sigma^2_{\phi_1}, \sigma^2_{\epsilon_i} \)

We define \( \theta^{(1:5)} = (\theta^1, \theta^2, \theta^3, \theta^4, \theta^5) \)

\[
\sigma^2_{\alpha} | \theta^{(1:5)}, y \sim IG \left( \frac{N + 2A_{\alpha}}{2}, \sum_{i=1}^{N} \alpha_i^2 + 2B_{\alpha} \right), \]

\[
\Omega^{-1}_{\delta} | \theta^{(1:5)}, y \sim Wish \left( [\delta \delta' + (\psi_{\delta} S_{\delta})]^{-1}, \psi_{\delta} + 1 \right), \]

\[
\sigma^2_{\phi_0} | \theta^{(1:5)}, y \sim IG \left( \frac{N + 2A_{\phi_0}}{2}, \sum_{i=1}^{N} (\phi_0 - \mu_{\phi_0})^2 + 2B_{\phi_0} \right), \]

\[
\sigma^2_{\phi_1} | \theta^{(1:5)}, y \sim IG \left( \frac{1 + 2A_{\phi_1}}{2}, \phi^2_1 + 2B_{\phi_1} \right), \]

\[
\sigma^2_{\epsilon_i} | \theta^{(1:5)}, y \sim IG \left( \frac{T_i + 2A_{\epsilon}}{2}, \frac{(r_i - \alpha_i - r_M \beta_i)^\prime (r_i - \alpha_i - r_M \beta_i) + 2B_{\epsilon}}{2} \right). \]

3.B Appendix II: Cross-Sectional Asset Pricing Tests

In this appendix we show how we account for measurement error in betas in the cross-sectional asset pricing tests. We consider the cross-sectional regression model described in section 3.6,

\[
r_{it} = \lambda_0 + \lambda_1 \beta_{it-1} + X'_{2t} x_{it-1} + v_{it}. \quad (3.18)\]
Conditional on \( \beta_{lt-1} \) we can estimate the parameters \( \lambda_t \), either Bayesian or classical, using the Fama-MacBeth approach. Let \( W_t = (1 \cdot \beta_t x_t)' \) and let \( \lambda_t \) denote the cross-sectional OLS estimator of \( \lambda_t \). The Fama-MacBeth estimator of the average \( \lambda \) is

\[
\hat{\lambda} = \frac{1}{T} \sum_{t} \hat{\lambda}_t = \frac{1}{T} (W_{t-1}' W_{t-1})^{-1} W_{t-1}' r_t,
\]

with covariance matrix

\[
S = Var \left( \sqrt{T}(\hat{\lambda} - \lambda) \right) = \frac{1}{T} \sum_{t} (\hat{\lambda}_t - \bar{\lambda})(\hat{\lambda}_t - \bar{\lambda})'.
\]

The Fama-MacBeth procedure overstates the precision of parameter estimates in the cross-sectional regressions because it ignores estimation errors in the \( \beta_{lt} \).

As explained in section 3.3, the Gibbs sampler has produced a series of \( L \) draws from the posterior density \( p(\beta | y) \), where \( \beta \) contains the entire collection of all \( \beta_{lt} \) and \( y \) is a shorthand for all data used in estimating the betas. Given the \( \beta_{lt}^{(i)} \) from the \( i^{th} \) iteration of the Gibbs sampler we can form the regressor matrix \( W_{lt}^{(i)} \) and using \( W_{lt}^{(i)} \) construct the conditional mean \( \hat{\lambda}^{(i)} \) and covariance matrix \( S^{(i)} \). From these we form the unconditional estimators

\[
\bar{\lambda} = \frac{1}{L} \sum_{i} \hat{\lambda}^{(i)}
\]

and

\[
\bar{S} = \frac{1}{L} \sum_{i} S^{(i)} + \frac{1}{L} \sum_{i} (\hat{\lambda}^{(i)} - \bar{\lambda})(\hat{\lambda}^{(i)} - \bar{\lambda})'.
\]

The estimates \( \bar{\lambda} \) and \( \bar{S} \) can be interpreted as the posterior mean and variance of the average \( \lambda_t \) if we assume that the prior on \( \lambda_t \) is uniform and \( \lambda \) does not affect the posterior density of \( \beta_{lt} \).
Chapter 4

Pension Funds: Performance, Benchmarks and Costs

Self-reporting biases and a deficiency of comprehensive return and cost data have severely hindered pension fund studies to date. The bias-free CEM data set resolves these issues and provides detailed information on fund-specific returns, costs and benchmarks for all types of pension plans and equity mandates. Using this data set we show that pension fund cost levels are substantially lower than mutual fund fees as a result of scale advantages and bargaining power by pension funds. We also document that benchmark-adjusted domestic net equity returns of US pension funds are very close to zero, which sharply contrasts the underperformance of mutual funds. We argue that the better monitoring capacity of pension funds enhances performance and also explains part of the performance difference with mutual funds.¹

4.1 Introduction

Pension funds are among the largest institutional investors in the US and a critical component of many people’s overall financial position. However, remarkably little is known about their performance and cost structures. By contrast, the mutual fund sector has been heavily scrutinized as have hedge funds. The lack of pension fund performance studies can be largely attributed to an absence of sufficient data which is a direct result of reporting guidelines. Mutual funds are required to report their performance and fees on a regular basis, whereas no such obligation exists for pension funds. The main contribution of this chapter is to employ the CEM pension fund database, which provides an excellent opportunity to evaluate pension fund performance.

¹ This chapter is based on the paper Bauer, Cremers, and Frehen (2009)
the performance of US pension funds and gives detailed insight in the variety of benchmarks used to evaluate asset managers and cost compositions of pension funds. The database covers approximately 40% of the US pension industry in terms of assets, representing a wide variety of fund sizes, plan types and equity mandates.

Pension fund return data in previous work was mostly of poor quality and often narrowed down to specific fund and plan types. In addition, every pension fund data set potentially suffered from biases as pension funds have no obligation to communicate their results. More importantly, most pension fund data are provided on a managed account level, i.e. returns are provided for managers employed by pension funds (see e.g. Beebower and Bergstrom (1977), Coggin, Fabozzi, and Rahman (1993) and Busse, Goyal, and Wahal (2009)). This reporting structure prohibits documentation of the performance of pension funds as such since managers are often employed by more than one pension fund and pension funds typically hire more than one manager. For those databases one typically does not know which manager is trading for which fund and therefore one cannot compute fund performance. Furthermore, data on benchmarks used by pension funds to evaluate their portfolio returns is to the best of our knowledge non-existent. A final drawback of data on pension funds is the lack of comprehensive cost data. Apart from a small number of surveys (e.g. McKinsey (2006) and Mercer (2006)), very little is known about the costs of pension funds.

The lack of comprehensive return, benchmark and cost data and possible biases have induced a broad diversity of conclusions on pension fund performance and costs. For example, Beebower and Bergstrom (1977) examine the performance of 148 US portfolios in a CAPM framework. In their study, the average portfolio outperforms the S&P 500 by 144 basis points per year. Coggin, Fabozzi, and Rahman (1993) document positive selectivity and negative timing skills for a random sample of 71 equity managers of US pension plans. Busse, Goyal, and Wahal (2009) perform the most complete study on pension fund accounts so far. They study 6,260 portfolios managed by institutional asset managers on behalf of defined benefit pension funds. Using a conditional multi-factor model, they find that the average fund manager outperforms the market by 124 basis points after expenses. By contrast, Brinson, Hood, and Beebower (1986) study 91 defined benefit plans and conclude that the funds underperform the S&P 500 by 110 basis points per year. Ippolito and Turner (1987) also document underperformance in a sample of 1,526 plans. They conclude that the S&P 500 on average has a return advantage of 44 basis points. Elton, Gruber, and Blake (2006) study mutual funds offered by defined contribution plans and show that they are beaten by the market by 31 basis points per year. In the most comprehensive study on plan level, Lakonishok, Shleifer, and Vishny (1992) show that 769 defined benefit plans lag the S&P 500 by 260 basis points yearly. Based on their figures it seems justified to question the future of the pension industry as they do.

The lack of consensus on pension fund performance is in marked contrast to the abundance of evidence on mutual fund underperformance. A majority of performance studies concludes that after expenses and trading costs, mutual funds
perform worse than a comparable passive market proxy. For example, Malkiel (1995) and Gruber (1996) observe that mutual funds on average underperform the market by the amount of expenses charged to investors. Chan, Chen, and Lakonishok (2002) corroborate the underperformance of the mutual fund industry in a study on mutual fund investment styles. More recently, Elton, Gruber, and Busse (2004) show that many index funds strongly underperform the S&P 500.

The above mentioned complications of pension fund data and contrasting findings of earlier pension fund studies illustrate how difficult it is to create a consistent and comprehensive picture of the performance of the US pension industry and give insight in the benchmarks and cost composition of this sector. The CEM data set enables us to address the aforementioned issues and strongly improve the performance and cost evaluation to date. We cover approximately 40% of the US pension sector in terms of assets under management and have annual data on pension plan level, i.e. returns, benchmarks and costs per fund. In particular, we examine the domestic equity performance of 463 defined benefit and 248 defined contribution pension plans and provide an overview of the benchmarks used by pension funds in their large and small cap mandates. Following Kenneth French, who first used the CEM data in his AFA presidential address on the cost of active investing (French, 2008), we use the data set to complete the picture on cost levels and their driving forces. By linking our data to other sources we are able to test for potential biases that result from the voluntary reporting, and ascertain that our data do not suffer from them. For most evaluations we differentiate between several mandate types i.e. large versus small capitalization stocks, actively versus passively managed portfolios and funds that are externally versus internally (only for defined benefit funds) managed.

We find that pension funds generate net domestic equity alphas very close to zero. Cost levels are substantially lower than in the mutual fund industry and range between 10 basis points for the largest defined benefit funds to 70 basis points for the smallest defined contribution funds. Heterogeneity in cost levels between funds is to a large extent driven by differences in fund size. Irrespective of fund size, externally and actively managed mandates are found be more expensive than their internally and passively managed counterparts.

A considerable majority of pension funds indicates that they benchmark their small cap equity mandates against the Russell 2000. Although all previous studies used the S&P 500 as a yardstick for performance, the CEM data show that this might not even be the right benchmark for the large cap mandates. The Wilshire 5000, Russel 1000 and Russell 3000 indices have conquered an equal market share in the large cap universe during the last decade.

The small but non-negative net domestic equity alphas of pension funds contrast sharply with the large amounts of evidence on mutual fund underperformance. Particularly, the lower cost levels of pension funds attribute to this gap in performance. We show that pension plan participants benefit from the size of their pension plans through lower cost levels in internally managed domestic equity portfolios. We also find that larger funds have lower cost levels in externally managed mandates, which is a clear proof of pension fund’s bargaining power with
external parties. Additionally, we provide evidence of the monitoring capacity of pension funds. More monitoring capacity leads to higher net alphas for large defined benefit funds, and to better performance in internally managed portfolios compared to externally managed ones. By contrast, Swensen (2005) shows that no such advantages exist for mutual fund investors. We thus argue that pension funds are not able to consistently outperform their benchmarks, but scale advantages in internally managed portfolios, bargaining power with external parties and monitoring abilities limit the magnitude and probability of underperformance. Investors in mutual funds on the other hand do not benefit from these advantages as a result of the for-profit character of the fund.

This chapter is organized as follows. Section 4.2 introduces our data set and measures possible biases. In section 4.3 we introduce the random coefficients panel model used to evaluate fund performance and the Fama-MacBeth methodology for measuring scale advantages in costs. In section 4.4 we give an overview of the most important benchmarks used to evaluate manager performance and study the impact of selecting different benchmarks in the evaluation procedure. In section 4.5 we measure the cost levels of various equity mandates and study the cost differences between funds. We evaluate fund performance in section 4.6 and put forward explanations for the performance difference between pension and mutual funds in section 4.7. Section 4.8 concludes.

### 4.2 Data

Defined benefit and defined contribution pension fund data are provided by CEM Benchmarking Inc. (CEM), which collects detailed information on pension fund costs, benchmarks and performance via yearly questionnaires. Table 4.1 illustrates the diversity and evolution of the equity database by reporting the number of funds for different fund types and the number of funds leaving the database for every sample year. Pension funds are grouped into corporate, public and other funds. The overview shows that the number of funds in the database is relatively stable, as is the relative share of corporate, public and other funds. It is therefore unlikely that our results will be driven by a single year, a single fund or a particular fund type. CEM collects data from multiple asset classes and numerous countries. In this chapter we will focus on domestic equity investments by US pension funds.

Table 4.1 also demonstrates the time frame of the analysis, 1990-2006 for defined benefit funds and 1997-2006 for defined contribution funds. A total of 463 defined benefit and 248 defined contribution funds report to CEM over the sample period. In any given year, approximately 150 defined benefit and 75 defined contribution US pension funds report. CEM retains information on funds entering and leaving the database. Additionally, the number of funds leaving the database is constant over time, further corroborating the fact that results will not be driven by individual funds or specific years and events.

The voluntary reporting to CEM makes the data vulnerable to biases. Table 4.1 gives a first insight in possible biases in our database. If funds would stop reporting
Table 4.1: Funds in the Database

This table reports the number of US pension funds incorporated in the defined benefit ('DB') and defined contribution ('DC') database in each year from 1990 to 2006. Panel A displays the number defined benefit funds in the sample period 1990-1996, Panel B the number of defined benefit funds in the sample period 1997-2006 and Panel C the number of defined contribution plans in the years 1997-2006. We report the total number of funds per year and split the total number up in corporate, public and other funds. The row labeled ('Exit') reports the number of funds leaving the database in that year. The CEM defined benefit database ranges from 1990 tot 2006 and the defined contribution database from 1997 to 2006. The last column ('Tot') reports the total number of funds in each category over the entire sample period.

**Panel A: DB 1990-1996**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>33</td>
<td>63</td>
<td>83</td>
<td>136</td>
<td>169</td>
<td>192</td>
<td>185</td>
</tr>
<tr>
<td>Corporate</td>
<td>25</td>
<td>39</td>
<td>55</td>
<td>84</td>
<td>101</td>
<td>106</td>
<td>112</td>
</tr>
<tr>
<td>Public</td>
<td>8</td>
<td>20</td>
<td>18</td>
<td>39</td>
<td>51</td>
<td>69</td>
<td>61</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>13</td>
<td>17</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Exit</td>
<td>0</td>
<td>11</td>
<td>17</td>
<td>19</td>
<td>35</td>
<td>39</td>
<td>43</td>
</tr>
</tbody>
</table>

**Panel B: DB 1997-2006**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>168</td>
<td>174</td>
<td>182</td>
<td>166</td>
<td>177</td>
<td>156</td>
<td>157</td>
<td>165</td>
<td>153</td>
<td>141</td>
<td>463</td>
</tr>
<tr>
<td>Corporate</td>
<td>95</td>
<td>92</td>
<td>86</td>
<td>74</td>
<td>81</td>
<td>68</td>
<td>72</td>
<td>77</td>
<td>68</td>
<td>62</td>
<td>271</td>
</tr>
<tr>
<td>Public</td>
<td>63</td>
<td>68</td>
<td>83</td>
<td>80</td>
<td>83</td>
<td>74</td>
<td>75</td>
<td>75</td>
<td>70</td>
<td>66</td>
<td>149</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>43</td>
</tr>
<tr>
<td>Exit</td>
<td>46</td>
<td>31</td>
<td>32</td>
<td>40</td>
<td>25</td>
<td>36</td>
<td>26</td>
<td>17</td>
<td>28</td>
<td>31</td>
<td>-</td>
</tr>
</tbody>
</table>

**Panel C: DC 1997-2006**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>61</td>
<td>71</td>
<td>65</td>
<td>67</td>
<td>85</td>
<td>74</td>
<td>89</td>
<td>86</td>
<td>91</td>
<td>118</td>
<td>248</td>
</tr>
<tr>
<td>Corporate</td>
<td>59</td>
<td>65</td>
<td>61</td>
<td>60</td>
<td>68</td>
<td>59</td>
<td>70</td>
<td>68</td>
<td>64</td>
<td>89</td>
<td>202</td>
</tr>
<tr>
<td>Public</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>14</td>
<td>12</td>
<td>17</td>
<td>16</td>
<td>23</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Exit</td>
<td>0</td>
<td>13</td>
<td>22</td>
<td>16</td>
<td>18</td>
<td>31</td>
<td>13</td>
<td>19</td>
<td>18</td>
<td>22</td>
<td>-</td>
</tr>
</tbody>
</table>
to CEM as a result of bad performance in a certain year, one would expect high numbers of exiting funds in years with bad market performance, e.g. 2001 and 2002. However, Table 4.1 gives no indication of high numbers of exiting funds in these years. We address the bias issue more formally by linking a secondary CEM data set to the Compustat SFAS 158 pension database. The Compustat database contains the yearly returns on pension assets\(^2\) (ROA) for all US corporate pension plans that have an obligation to report to the SEC from 1998 onwards. Hence, the Compustat data set does not suffer from biases related to voluntary reporting.

In addition to the aforementioned, primary data set (containing returns, benchmarks and costs for anonymous pension funds), CEM provided us with a secondary database containing fund names and entry and exit years in the primary data set. However, this secondary database cannot be linked with the primary, anonymous data source. We only matched the secondary database to the Compustat data, thus respecting the anonymity of the funds in the primary set containing performance, benchmark and cost figures. The matching procedure\(^3\) resulted in respectively 67% and 49% matched defined benefit and defined contribution pension funds in the CEM database.

We first test for biases by categorizing the matched Compustat ROA observations into two groups. The first group contains ROAs of years in which funds have decided to stop reporting, i.e. the first year that they did not report to CEM anymore. The second group contains all remaining data points. Consequently, funds that have decided to cease reporting have observations in both groups. Funds that report the entire period only appear in group 2. By comparing the means of the groups, we test whether funds decide to stop reporting as a result of bad or good performance.

Thereafter we repeat this test for the entering years, i.e. we categorize the matched Compustat data points into two groups. Group 1 contains all ROAs of years in which funds decide to report to CEM for the first time (excluding the first year of the data set). The second group contains all the other data points. Again, we test if there is a significant difference in mean ROA between the two groups.

Panels A and B of Table 4.2 show that there is no evidence of a self-reporting bias in the exiting or entering years. In Panel A, we report the difference in the mean ROA between the above defined groups, i.e. we measure the difference in mean ROA of exiting year data points and all remaining data points. In Panel B, we repeat this procedure for entering years. The differences in mean ROA are all insignificant. Since most funds have observations in both groups and irregularly enter and exit the database (see Table 4.1), it is not necessary to adjust for autocorrelation or heteroskedasticity in this test.

In a second bias test, we conduct a logit regression of a dummy variable indicating presence in the CEM database on the Compustat ROA performance for

---

\(^2\) This is the total return over all asset classes (contrasting the equity focus in the rest of the chapter). We argue that funds are likely to stop reporting if their overall performance is bad. This decision does not merely depend on their equity performance.

\(^3\) Funds are matched based on their names. Since some pension funds in the CEM database have no obligation to report to the SEC (e.g. because they are too small) it is impossible to match 100% of the funds. Name changes are not considered as entries or exits.
Table 4.2: Sample Mean Bias Measurement

This table reports the results of two tests that measure possible biases in the CEM database. Panel A reports the difference in group means of two groups of data points and its corresponding p-value. One group contains total returns on pension assets from the Compustat SFAS 158 pension database in years that funds decided to stop ('Exit') reporting to CEM. The other group of data points consists of total returns on pension assets in all other years. We perform the mean comparison for both defined benefit ('DB') and defined contribution ('DC') pension funds after matching the Compustat and CEM database. Mean returns in exiting years are subtracted from the remaining returns. In the matching procedure 114 of the 171 CEM corporate defined benefit funds and 93 of the 191 corporate defined contribution funds were matched with the Compustat database. In Panel B, we employ the same test, but divide the data points in years in which funds enter the data set ('Entry') and all remaining observations. Panel C displays the parameter estimates of a logit regression ('Logit') of a dummy variable indicating presence in the CEM database on Compustat ROA in that year, together with the corresponding p-values.

<table>
<thead>
<tr>
<th>Panel A: Exit</th>
<th>DB</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.80%</td>
<td>2.10%</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.31)</td>
<td>(0.09)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Entry</th>
<th>DB</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.10%</td>
<td>-0.10%</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.40)</td>
<td>(0.40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Logit</th>
<th>DB</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>-0.46</td>
<td>-0.32</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>ROA</td>
<td>0.36</td>
<td>0.49</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.53)</td>
<td>(0.43)</td>
</tr>
</tbody>
</table>
the matched sample. The corresponding p-values in panel C of Table 4.2 again provide no evidence of biases in the CEM database. The high degree of freedom ensures sufficient power to perform this test.

The unique structure of the CEM database allows for an accurate evaluation of the cost and performance of domestic equity investments by US pension funds. It provides the opportunity to evaluate large and small cap, actively and passively managed and internally and externally managed (only for defined benefit funds) equity investments separately. Figure 4.1 shows the structure of our pension fund database by representing all equity classifications. When starting at the highest aggregation level, containing all equity investments, subsequent breakdowns create different aggregation levels as indicated in Figure 4.1. Data provided by CEM are reported on low aggregation levels (e.g. gross returns on internally, passively managed large cap stocks). For this reason, we are able to measure differences between investment styles. In order to measure these differentials, we need to aggregate the data to higher levels. Higher aggregation-level domestic equity returns are computed as value-weighted averages of lower level returns with lower level holdings as weights. Holdings are aggregated by addition of lower-level holdings. If, on a certain level, funds report a positive holding but no return, value-weighted aggregation is not possible. We exclude these observations. Performance is measured net of costs (and in some cases also net of fund-specific benchmarks), and thus consists of the difference of variables. If two observations need to be added or subtracted and one of them is missing, we regard the sum or difference as missing as well.

In Table 4.3, we report the portfolio composition after splitting the equity holdings into various investment classifications. Generally, defined benefit funds have a major part invested in large cap mandates (90%). Defined benefit funds employ more active (70%) than passive managers and most mandates are outsourced to external managers (87%). These figures are fairly consistent for corporate, public and other funds in the database. However, when we focus on the 30% largest and smallest portfolio holdings we detect fundamentally different equity portfolio compositions. Larger sized equity portfolios tend to give more relative weight to large cap and passively managed mandates. Also the larger portfolios are more often internally managed. The relatively large weight in large cap mandates might reflect possible diseconomies of scale: as a result of the sheer size of the total equity holdings it is more complicated to invest large amounts in small cap mandates. Higher percentages of internal management can easily be explained by economies of scale; the larger the fund, the lower the relative costs of internal management. For defined contribution plans, the differences in equity holdings between large and small portfolios are less pronounced, since the size differences are smaller than in the defined benefit universe. The patterns in defined contribution equity holdings are comparable to the patterns in small defined benefit plans, which are of similar fund size.
Figure 4.1: Data Structure

This figure displays the structure of the CEM database. CEM provides data on aggregation levels 1 and 2 and we aggregate data up to the highest level. Mandates are split up into large and small cap and actively and passively managed. For defined benefit plans, CEM also differentiates between internally and externally managed portfolios.
Table 4.3: Equity Holdings per Classification

In this table we report the percentages invested in large cap versus small cap, actively managed versus passively managed mandates, and internally versus externally (only for defined benefit funds) managed domestic US equity mandates. Panel A displays results for defined benefit ('DB') plans from 1990 to 2006. Panel B contains percentages of equity holdings for each classification in the defined contribution ('DC') universe, for the sample period 1997 to 2006. We report results for all funds in the database ('Tot'), all corporate ('Cor'), public ('Pub') and other ('Oth') funds. Moreover, we provide information for the 30% largest ('L30%') and 30% smallest ('S30%') funds in terms of domestic equity holdings.

**Panel A: DB**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Cor</th>
<th>Pub</th>
<th>Oth</th>
<th>L30%</th>
<th>S30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large cap</td>
<td>90.49</td>
<td>90.90</td>
<td>89.87</td>
<td>90.26</td>
<td>93.01</td>
<td>88.30</td>
</tr>
<tr>
<td>Small cap</td>
<td>9.51</td>
<td>9.10</td>
<td>10.13</td>
<td>9.74</td>
<td>6.99</td>
<td>11.70</td>
</tr>
<tr>
<td>Active</td>
<td>69.41</td>
<td>73.66</td>
<td>63.30</td>
<td>68.61</td>
<td>57.24</td>
<td>78.87</td>
</tr>
<tr>
<td>Passive</td>
<td>30.59</td>
<td>26.34</td>
<td>36.70</td>
<td>31.39</td>
<td>42.76</td>
<td>21.13</td>
</tr>
<tr>
<td>Internal</td>
<td>13.84</td>
<td>9.62</td>
<td>20.15</td>
<td>7.13</td>
<td>27.74</td>
<td>2.86</td>
</tr>
<tr>
<td>External</td>
<td>87.10</td>
<td>90.38</td>
<td>79.85</td>
<td>92.87</td>
<td>72.26</td>
<td>97.14</td>
</tr>
</tbody>
</table>

**Panel B: DC**

<table>
<thead>
<tr>
<th></th>
<th>Tot</th>
<th>Cor</th>
<th>Pub</th>
<th>Oth</th>
<th>L30%</th>
<th>S30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large cap</td>
<td>83.33</td>
<td>83.56</td>
<td>84.62</td>
<td>78.76</td>
<td>84.72</td>
<td>81.37</td>
</tr>
<tr>
<td>Active</td>
<td>55.07</td>
<td>54.26</td>
<td>70.37</td>
<td>29.97</td>
<td>52.70</td>
<td>55.11</td>
</tr>
<tr>
<td>Passive</td>
<td>44.93</td>
<td>45.74</td>
<td>29.63</td>
<td>70.03</td>
<td>47.30</td>
<td>44.89</td>
</tr>
</tbody>
</table>
4.3 Random Coefficients Panel Model

In this section, we introduce the panel data technique used to evaluate pension fund performance and explain the Fama-MacBeth methodology adopted to study size effects in costs. Due to the yearly frequency of our data set, standard time-series regression methods cannot be employed. We therefore rely on a random coefficients model to test hypotheses on fund performance. Our panel methodology described in this section differs from pooled panel regressions by allowing the alphas and betas to vary over the cross-section instead of assuming common parameters for all funds. Our assumption of a common distribution for alphas and betas is less restrictive than common parameters across funds. The methodology also differs from Timmermann, Blake, Tonks, and Wermers (2009), by allowing idiosyncratic risk to vary cross-sectionally. As a direct consequence of our generalizations we can estimate our model only for funds with sufficient observations. We measure the cross-sectional explanatory power of fund size for fund-specific costs using the well-known Fama-Macbeth methodology, (Fama and MacBeth, 1973).

Standard performance evaluation studies regress fund returns on Fama-French factors and measure the performance by fund-specific alphas. However, running time series regressions with unbalanced annual data from 1990 to 2006 either leads to extremely noisy estimates or is even impossible as a result of insufficient data points. Since this chapter focuses on the performance of the pension sector rather than individual fund performance, there is no need to estimate fund-specific alphas directly and we can rely on a hierarchical panel model for which we estimate higher order parameters which can be interpreted as pension fund industry alphas and betas. The random coefficients model assumes that fund-specific alphas and betas of the classical Fama-French factors are independently drawn from hierarchical distributions with common parameters. The hierarchical parameters can be estimated with high precision as we benefit from the large cross-section of our sample. We define net excess returns \( R_{it,e} \) as gross domestic equity returns net of costs and 3 month US T-Bill rate \( R_f \) and characterize our model as

\[
R_{it,e} = \alpha_i + \beta_i' FF_t + \eta_{it}, \tag{4.1}
\]

where we assume that \( \alpha_i \) and \( \beta_i \) are drawn independently from distributions with constant means and variances, \( FF_t \) the time \( t \) Fama-French factors \( (R_M - R_f, SMB, HML \text{ and Momentum}) \) and \( \eta_{it} \) a normally distributed error term with zero mean,

\[
\alpha_i \sim N \left( \alpha, \sigma^2_\alpha \right), \quad \beta_i \sim N \left( \beta, \Omega_\beta \right), \quad \eta_{it} \sim N \left( 0, \sigma^2_{\eta_{it}} \right),
\]

where we assume that \( \Omega_\beta \) is a diagonal matrix and \( \beta \) is a vector with factor loadings \( (\beta_M, \beta_{SMB}, \beta_{HML} \text{ and } \beta_{MOM}) \). We also regress net benchmark adjusted returns \( (R_{it} - BM R_{it}) \) on the same set of factors. This model can equivalently be described as

\[
R_{it} - BM R_{it} = \alpha_i + \beta_i' FF_t + \nu_{it}, \tag{4.2}
\]
with \( R_{it} \) the time \( t \) return on domestic equity investments after subtraction of costs, \( BMR_{it} \) the time \( t \) benchmark return, \( \nu_{it} \) the zero-mean error term and all other parameters and factors equivalent to equation (4.1).

We augment the number of factors in equation (4.1) to study the effect of fund size on the risk-adjusted performance. This extended version of our random coefficients model is described as

\[
R_{it,e} = \alpha_{0i} + \alpha_{1i} \log(\text{Size}_{it}) + \beta_i' FF_t + \xi_{it},
\]

where we again assume that alphas and betas are independently drawn from distributions

\[
\alpha_{0i} \sim N(\alpha_0, \sigma_{\alpha_0}^2), \quad \alpha_{1i} \sim N(\alpha_1, \sigma_{\alpha_1}^2), \quad \beta_i \sim N(\beta, \Omega_{\beta}), \quad \xi_{it} \sim N(0, \sigma_{\xi}^2),
\]

with \( \Omega_{\beta} \) a diagonal matrix. For a more detailed description of our random coefficients model, we refer to Swamy (1970).

In addition to performance measurement, we test cross-sectionally for scale advantages in fund costs by regressing fund-specific costs defined as \( C_{it} \) on a constant and log fund size in million US dollar,

\[
C_{it} = a_{0t} + a_{1t} \log(\text{Size}_{it}) + \epsilon_{it},
\]

with \( \epsilon_{it} \) a normally distributed zero-mean error term. Thereafter we regress the parameter estimates from the cross-sectional regressions in a time series regression on a constant, and correct for autocorrelation and heteroskedasticity using the Newey-West procedure.

### 4.4 The Importance of Choosing the Right Market Factor

In this section we give an overview of the benchmarks used by pension funds to evaluate performance and show the impact on the performance figures of choosing different market factors. As summarized in the introduction of this chapter, previous literature measured pension fund performance against the S&P 500. However to date no study has given any insights in the benchmarks actually used by pension funds to evaluate their asset managers. For evaluation purposes it is interesting for pension funds to know which benchmarks are popular among their peers. Besides, knowing the right benchmarks is crucial in standard risk-adjustment procedures for scholars as choosing the wrong market factors may lead to severe biases in fund alphas and betas.

#### 4.4.1 Which Benchmarks do Pension Funds use?

In their yearly questionnaires, CEM requests pension funds to report the benchmarks used to evaluate large and small cap mandates. Funds are required to provide an exact description of the benchmark in combination with the corresponding
4.4 The Importance of Choosing the Right Market Factor

Figure 4.2: Large Cap Benchmarks

This figure displays the percentage of defined benefit pension funds which uses a specific benchmark in a certain year. We show the percentage of funds which use the S&P 500, Russell 1000, Russell 3000 or Wilshire 4500 as benchmark for their large cap equity mandates from 1990 to 2006.

return. Consequently, we can construct market shares for each benchmark, for every given year. Figure 4.2 gives an overview of each benchmark’s relative market share in large cap mandates\(^4\). Figure 4.3 displays the evolution of small cap benchmark market shares from 1990 to 2006.

Figure 4.2 shows that the S&P 500 has been the most important large cap benchmark throughout the nineties. However, nowadays many funds also use the Russell 1000, Russell 3000 and Wilshire 5000 as yardstick for their returns on large capitalization stocks. Figure 4.3 shows that the Russell 2000 is the dominant small cap benchmark throughout the entire sample period. Combining these two findings we can conclude that it is not only important to differentiate between large and small cap mandates when evaluating pension fund performance, the benchmarks

\(^4\) There are two reasons why percentages market shares do not sum to 100. First, we omitted several benchmarks with small market shares for representation purposes. Second, pension funds sometimes report customized benchmarks which are a linear combination of standard benchmarks, e.g. 50\%*S&P 500 + 50\%*Russell 3000. We also omitted the customized benchmarks from figures 4.2 and 4.3.
Figure 4.3: **Small Cap Benchmarks**

This figure displays the percentage of defined benefit pension funds which uses a specific benchmark in a certain year. We show the percentage of funds which use the Russell 2000, Russell 2500, Russell 3000, Wilshire 4500, Wilshire 5000, S&P 400, S&P 500 or S&P 600 as benchmark for their small cap equity mandates from 1990 to 2006.
4.4 The Importance of Choosing the Right Market Factor

4.4.2 Why Choosing the Right Market Factor is Important

Having concluded that there is broad cross-sectional and time-series variation in benchmarks used by pension funds to evaluate their asset managers, the natural questions arises: What is the effect of using different market factors (or benchmarks) for performance evaluation by scholars? In this subsection we study the effect of using different market factors in the risk-adjustment procedure described in section 4.3.

Table 4.4: Benchmark Effects Performance

This table reports the parameter estimates of a random coefficients panel regression of total net excess returns of domestic US equity mandates ('R_{it,e}') of defined benefit plans, on excess market return ('R_{m,t}'), Size ('SMB'), Book-to-Market ('BTM') and Momentum ('MOM'), for the sample period 1990-2006. The random coefficients model can be characterized as

\[ R_{it,e} = \alpha_i + \beta_{R_{m,t}} R_{m,t} + \beta_{SMB} SMB_t + \beta_{BTM} BTM_t + \beta_{MOM} MOM_t + \eta_{it}, \]

with normally distributed mean-zero errors. We assume that the parameters \( \alpha_i \), \( \beta_{R_{m,t}} \), \( \beta_{SMB} \), \( \beta_{BTM} \) and \( \beta_{MOM} \) are drawn independently from hierarchical distributions \( N(\alpha, \sigma^2_{\alpha}) \), \( N(\beta_{R_{m,t}}, \sigma^2_{\beta_{R_{m,t}}}) \), \( N(\beta_{SMB}, \sigma^2_{\beta_{SMB}}) \), \( N(\beta_{BTM}, \sigma^2_{\beta_{BTM}}) \) and \( N(\beta_{MOM}, \sigma^2_{\beta_{MOM}}) \). This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different market factor variables. In the first row we use the Fama-French market factor, in the second the Russell 3000 and in the last row the S&P 500 factor.

<table>
<thead>
<tr>
<th></th>
<th>( \alpha )</th>
<th>( \beta_{R_{m,t}} )</th>
<th>( \beta_{SMB} )</th>
<th>( \beta_{BTM} )</th>
<th>( \beta_{MOM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>0.0103</td>
<td>0.9289</td>
<td>-0.0165</td>
<td>0.0039</td>
<td>-0.0727</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.542)</td>
<td>(0.861)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Russell 3000</td>
<td>0.0048</td>
<td>0.9456</td>
<td>0.0522</td>
<td>-0.0108</td>
<td>-0.0199</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.084)</td>
<td>(0.000)</td>
<td>(0.052)</td>
<td>(0.617)</td>
<td>(0.337)</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.0041</td>
<td>0.9358</td>
<td>0.1259</td>
<td>-0.0098</td>
<td>-0.0274</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.194)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.685)</td>
<td>(0.230)</td>
</tr>
</tbody>
</table>

In Table 4.4 we report the parameter estimates of the random coefficients panel regression of annual net US equity returns on excess market return, Size, Book-to-Market and Momentum. We investigate the choice of three different market proxies: the excess market return as constructed by Kenneth French\(^5\), the Russell 3000 and the S&P500. In the first row of Table 4.4, we show that alpha is more than 1% per year and significant at a 1% level, when using the FF market factor. The average fund has a negative loading on the SMB and a positive one on the BTM factor, indicating a tilt towards large value stocks. However, when we replace the

FF market factor by either the Russell 3000 or the S&P 500, alphas decrease (below 1%) and become insignificant, $\beta_{Rm}$ increases slightly, and the negative loading on the SMB factor switches to positive while the BTM-beta decreases and becomes negative. This corroborates the finding by Cremers and Petajisto (2009) that the FF market factor puts a disproportionately large weight on small growth stocks. Table 4.4 also indicates that the net equity returns of US pension funds are better explained by the S&P 500 index or the Russell 3000 index than by the FF market factor. We proceed our empirical analysis by using the S&P 500 as market factor in our risk-adjustment procedure, which results in conservative estimates of fund alphas (being the lowest in Table 4.4). To determine how robust our results are with respect to benchmark choices we will also study benchmark-adjusted returns in our performance evaluation.

### 4.5 Pension Fund Cost Levels and Drivers

The CEM database contains detailed and comprehensive information on the costs of US pension funds. In this section, we provide an overview of the general level of the costs of domestic equity investments, the differences in costs between various plan and mandate types and the drivers of these differences.

Table 4.5 shows that the average cost level for US equity portfolios is roughly 30 basis points for a defined benefit pension plan and 50 basis points for a defined contribution plan. We also document that cost levels of pension funds range between 10 basis points for the largest defined benefit funds to 70 basis points for the smallest defined contribution plans. Swensen (2005) shows that average mutual fund fees amount to 150 basis points for both load and no-load funds. We thus conclude that average costs of pension plans are substantially lower than mutual fund fees. Table 4.5 also indicates that larger mandates have lower cost levels. Many of the differences (public vs. corporate and defined benefit vs. defined contribution) in Table 4.5 may therefore be explained by size discrepancies. Evidence of scale advantages in costs is most clear when we compute cost levels for the 30% largest and 30% smallest funds.

Table 4.6 displays cost levels for various mandate types and clearly indicates that size is not the only driver of pension fund costs. Outsourced portfolios are more expensive than internally managed ones. Likewise, actively managed mandates have higher cost levels than their passively managed counterparts. Even if we distinguish between larger and smaller funds these cost differences remain. We thus conclude that besides fund size, mandate types also have explanatory power for cost levels.

Tables 4.5 and 4.6 have shown that fund size is an important driver of costs for domestic equity investments. Both tables provide evidence of substantial scale advantages in costs. In Table 4.7 we test more formally whether large pension funds benefit from reduced cost levels. Using the Fama-MacBeth methodology described in section 4.3, we regress costs for domestic equity investments on log fund size. For exposition purposes we have multiplied cost levels with 100.
### 4.5 Pension Fund Cost Levels and Drivers

This table provides information on the costs of domestic equity investments ('Costs') and holdings size of domestic equity investments ('Size') by US pension funds. Costs levels are displayed in basis points and size in millions of US dollar. We provide information on defined benefit ('DB') pension plans for the sample period 1990-2006 and defined contribution ('DC') pension plans for the sample period 1997-2006. We show results for the full sample, and for corporate, public and other pension funds separately. All costs numbers are calculated as follows: first the equally weighted average per year is calculated. Then the time-series mean for the corresponding sample period is calculated. Additionally, we yearly sort all funds on size of the holdings and make a group of largest 30% and smallest 30% of funds.

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DB</td>
<td>DC</td>
<td>DB</td>
<td>DC</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28.75</td>
<td>51.09</td>
<td>4,183</td>
<td>1,186</td>
<td></td>
</tr>
<tr>
<td>Largest 30%</td>
<td>16.09</td>
<td>42.47</td>
<td>10,759</td>
<td>3,023</td>
<td></td>
</tr>
<tr>
<td>Smallest 30%</td>
<td>39.92</td>
<td>61.30</td>
<td>389</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>Corporate</td>
<td>32.63</td>
<td>48.45</td>
<td>2,260</td>
<td>1,190</td>
<td></td>
</tr>
<tr>
<td>Largest 30%</td>
<td>24.42</td>
<td>39.16</td>
<td>5,860</td>
<td>2,695</td>
<td></td>
</tr>
<tr>
<td>Smallest 30%</td>
<td>40.22</td>
<td>56.90</td>
<td>370</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>21.33</td>
<td>58.68</td>
<td>6,964</td>
<td>1,629</td>
<td></td>
</tr>
<tr>
<td>Largest 30%</td>
<td>9.32</td>
<td>55.26</td>
<td>17,313</td>
<td>4,387</td>
<td></td>
</tr>
<tr>
<td>Smallest 30%</td>
<td>34.05</td>
<td>73.15</td>
<td>661</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>29.46</td>
<td>-</td>
<td>2,863</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Largest 30%</td>
<td>14.75</td>
<td>-</td>
<td>6,275</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Smallest 30%</td>
<td>42.75</td>
<td>-</td>
<td>291</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.6: Costs per Equity Classification

In this table we report costs in basis points per mandate type of domestic equity investments by defined benefit ('DB') and defined contribution ('DC') US pension funds in their respective sample periods, i.e. 1990-2006 for defined benefit and 1997-2006 for defined contribution plans. We distinguish between large cap, small cap, actively and passively managed mandates, and we also differentiate between costs for internally and externally managed mandates for defined benefits plans only. All results are displayed for the entire sample and subsamples of the largest and smallest 30% of the funds.

<table>
<thead>
<tr>
<th>Investment Type</th>
<th>DB</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large cap</td>
<td>27.36</td>
<td>46.21</td>
</tr>
<tr>
<td>Large cap Largest 30%</td>
<td>16.54</td>
<td>38.29</td>
</tr>
<tr>
<td>Large cap Smallest 30%</td>
<td>36.79</td>
<td>54.40</td>
</tr>
<tr>
<td>Small cap</td>
<td>42.58</td>
<td>76.88</td>
</tr>
<tr>
<td>Small cap Largest 30%</td>
<td>32.63</td>
<td>66.90</td>
</tr>
<tr>
<td>Small cap Smallest 30%</td>
<td>60.57</td>
<td>91.49</td>
</tr>
<tr>
<td>Active</td>
<td>41.72</td>
<td>76.46</td>
</tr>
<tr>
<td>Active Largest 30%</td>
<td>31.22</td>
<td>70.56</td>
</tr>
<tr>
<td>Active Smallest 30%</td>
<td>50.69</td>
<td>86.32</td>
</tr>
<tr>
<td>Passive</td>
<td>5.24</td>
<td>20.75</td>
</tr>
<tr>
<td>Passive Largest 30%</td>
<td>2.92</td>
<td>13.49</td>
</tr>
<tr>
<td>Passive Smallest 30%</td>
<td>9.09</td>
<td>28.93</td>
</tr>
<tr>
<td>Internal</td>
<td>9.84</td>
<td>-</td>
</tr>
<tr>
<td>Internal Largest 30%</td>
<td>7.12</td>
<td>-</td>
</tr>
<tr>
<td>Internal Smallest 30%</td>
<td>8.84</td>
<td>-</td>
</tr>
<tr>
<td>External</td>
<td>43.61</td>
<td>-</td>
</tr>
<tr>
<td>External Largest 30%</td>
<td>33.11</td>
<td>-</td>
</tr>
<tr>
<td>External Smallest 30%</td>
<td>52.39</td>
<td>-</td>
</tr>
</tbody>
</table>
4.5 Pension Fund Cost Levels and Drivers

Table 4.7: Size Effects in Costs

This table reports the parameter estimates of Fama-MacBeth regressions of costs of domestic equity investments by US defined benefit ('DB') and defined contribution ('DC') plans, on a constant and log fund size. In the first stage we cross-sectionally regress fund-specific costs on a constant and log fund size in million US dollar for every year in our sample period, i.e. 1990-2006 for DB plans and 1997-2006 for defined contribution plans. Thereafter we regress the constants and loadings in a time series regression on a constant, correcting the standard errors via the Newey-West procedure. Total costs of domestic US equity investments are the dependent variable in the first row. We repeat the procedure for the costs of large cap, small cap, actively managed, passively managed, internally managed and externally managed mandates. For exposition purposes we have multiplied cost levels with 100 and therefore represent the figures in basis points.

<table>
<thead>
<tr>
<th></th>
<th>DB</th>
<th></th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_0$</td>
<td>$a_{Size}$</td>
<td>$a_0$</td>
</tr>
<tr>
<td>Total</td>
<td>112.2244</td>
<td>-9.2404</td>
<td>56.7527</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Large cap</td>
<td>93.1465</td>
<td>-7.8851</td>
<td>95.2824</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Small cap</td>
<td>87.8018</td>
<td>-8.4283</td>
<td>152.8869</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Active</td>
<td>114.6403</td>
<td>-8.4283</td>
<td>216.7413</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Passive</td>
<td>28.6734</td>
<td>-2.5291</td>
<td>140.6787</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>External</td>
<td>102.6544</td>
<td>-7.6933</td>
<td>-</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>-</td>
</tr>
<tr>
<td>Internal</td>
<td>23.7273</td>
<td>-1.7792</td>
<td>-</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>-</td>
</tr>
</tbody>
</table>
4 Pension Funds: Performance, Benchmarks and Costs

The negative loading on log fund size for all mandate types implies that larger pension funds have lower cost levels than smaller ones. Although parameter estimates do not turn up to be significant in all classifications, the pattern of negative loadings is clear and unambiguous. Pension funds display considerable scale advantages in their costs of domestic equity investments. If we zoom in on externally managed mandates the negative loading is even more remarkable since this is not caused by cost reductions within the fund. The significant negative loading indicates that large funds have more bargaining power and can therefore reduce costs of external portfolio management.

From an economic point of view, one may expect larger pension funds to have lower cost levels since they can divide their fixed costs over more participants. Nevertheless Swensen (2005) and Carhart (1997) have shown that as a result of profit objectives in the mutual fund industry, no scale advantages are documented in that sector. We thus conclude that pension plan participants benefit from scale advantages, whereas in the mutual fund sector the fund takes advantage of the larger size.

4.6 Risk-adjusted Performance

We measure the risk-adjusted domestic equity performance of US pension funds. First, we specify performance before benchmark-adjustments (net excess returns) as gross domestic equity returns minus the risk free rate and fund-specific costs. Net benchmark-adjusted returns are acquired by subtracting fund-specific benchmark returns and costs from gross domestic equity returns. We regress the benchmark-adjusted returns on a constant using our standard random coefficients model. Thereafter, we add risk factors (Size, BTM and Momentum) as explanatory variables, to pick up outperformance due to excessive risk-taking. A possible drawback of this simple risk-adjustment methodology is the strategic choice of benchmarks. Managers may change performance benchmarks in such a way that outperformance and consequently bonus payments are easier to attain. We therefore also regress net excess returns on the S&P 500, Size, BTM and Momentum and measure industry performance via the alphas of the common distributions, which we interpret as pension fund industry alphas (see section 4.3 for more details). Hence if managers outperform the market by simple small cap, value or momentum strategies we again pick up these effects.

Table 4.8 displays parameter estimates of a random coefficients panel regression of benchmark-adjusted returns on a single constant and on a constant augmented with the Fama-French factors for defined benefit pension plans, as described in equation (4.2). Table 4.9 shows the same results for defined contribution plans.

The mean benchmark-adjusted returns reported in the final columns (’c’), clearly show that pension funds perform very close to their benchmarks. We split the universe in large and small cap, actively and passively managed and internally and externally managed mandates and document that benchmark-adjusted returns do not differ significantly from zero except for actively managed mandates.
Table 4.8: Benchmark Adjusted Performance Defined Benefit Plans

This table reports the parameter estimates of a random coefficients panel regression of fund net benchmark adjusted domestic equity returns (\( R_{it} - BMR_{it} \)) of defined benefit US pension plans on excess S&P 500 return (\( R_m \)), Size (\( SMB \)), Book-to-Market (\( BTM \)) and Momentum (\( MOM \)), for the sample period 1990-2006. The random coefficients model can be characterized as

\[
R_{it} - BMR_{it} = \alpha_i + \beta_{R_m,i}R_m + \beta_{SMB,i}SMB + \beta_{BTM,i}BTM + \beta_{MOM,i}MOM + \eta_{it},
\]

with normally distributed mean-zero errors. We assume that the parameters \( \alpha_i, \beta_{R_m,i}, \beta_{SMB,i}, \beta_{BTM,i} \) and \( \beta_{MOM,i} \) are drawn independently from hierarchical distributions \( N(\alpha, \sigma^2_\alpha), N(\beta_{R_m}, \sigma^2_{\beta_{R_m}}), N(\beta_{SMB}, \sigma^2_{\beta_{SMB}}), N(\beta_{BTM}, \sigma^2_{\beta_{BTM}}) \) and \( N(\beta_{MOM}, \sigma^2_{\beta_{MOM}}) \). This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different dependent variables. Total net benchmark-adjusted returns of domestic US equity investments are the dependent variable in the first row. Likewise, we regress the net benchmark-adjusted returns of large cap, small cap, actively managed, passively managed, internally managed and externally managed mandates on the same factors. The last column displays the results of a regression of the same dependent variables on a constant, with its corresponding p-value.

<table>
<thead>
<tr>
<th></th>
<th>( \alpha )</th>
<th>( \beta_{R_m} )</th>
<th>( \beta_{SMB} )</th>
<th>( \beta_{BTM} )</th>
<th>( \beta_{MOM} )</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.0043</td>
<td>-0.0456</td>
<td>0.0367</td>
<td>-0.0237</td>
<td>-0.0110</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.000)</td>
<td>(0.152)</td>
<td>(0.298)</td>
<td>(0.546)</td>
<td>(0.757)</td>
</tr>
<tr>
<td>Large cap</td>
<td>0.0018</td>
<td>-0.0422</td>
<td>0.0321</td>
<td>0.0005</td>
<td>-0.0143</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.515)</td>
<td>(0.000)</td>
<td>(0.222)</td>
<td>(0.984)</td>
<td>(0.487)</td>
<td>(0.306)</td>
</tr>
<tr>
<td>Small cap</td>
<td>0.0343</td>
<td>-0.1027</td>
<td>0.0374</td>
<td>-0.2486</td>
<td>-0.0917</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.274)</td>
<td>(0.000)</td>
<td>(0.120)</td>
<td>(0.616)</td>
</tr>
<tr>
<td>Active</td>
<td>0.0045</td>
<td>-0.0605</td>
<td>0.0546</td>
<td>-0.0213</td>
<td>-0.0112</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.000)</td>
<td>(0.057)</td>
<td>(0.441)</td>
<td>(0.604)</td>
<td>(0.435)</td>
</tr>
<tr>
<td>Passive</td>
<td>0.0043</td>
<td>-0.0230</td>
<td>-0.0077</td>
<td>-0.0208</td>
<td>-0.0160</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.057)</td>
<td>(0.582)</td>
<td>(0.296)</td>
<td>(0.370)</td>
<td>(0.318)</td>
</tr>
<tr>
<td>Internal</td>
<td>0.0156</td>
<td>-0.0927</td>
<td>0.0527</td>
<td>-0.0269</td>
<td>-0.0282</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.112)</td>
<td>(0.440)</td>
<td>(0.722)</td>
<td>(0.811)</td>
<td>(0.574)</td>
</tr>
<tr>
<td>External</td>
<td>0.0045</td>
<td>-0.0484</td>
<td>0.0395</td>
<td>-0.0305</td>
<td>-0.0103</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.000)</td>
<td>(0.136)</td>
<td>(0.204)</td>
<td>(0.601)</td>
<td>(0.511)</td>
</tr>
</tbody>
</table>
Table 4.9: Benchmark Adjusted Performance Defined Contribution Plans

This table reports the parameter estimates of a random coefficients panel regression of net benchmark adjusted domestic equity returns ($R_{it} - BMR_{it}$) of defined contribution US pension plans on excess S&P 500 return ($R_m$), Size ($SMB$), Book-to-Market ($BTM$) and Momentum ($MOM$), for the sample period 1997-2006. The random coefficients model can be characterized as

$$R_{it} - BMR_{it} = \alpha_i + \beta_{R_m,i}R_m + \beta_{SMB,i}SMB + \beta_{BTM,i}BTM + \beta_{MOM,i}MOM + \epsilon_{it},$$

with normally distributed mean-zero errors. We assume that the parameters $\alpha_i$, $\beta_{R_m,i}$, $\beta_{SMB,i}$, $\beta_{BTM,i}$ and $\beta_{MOM,i}$ are drawn independently, from hierarchical distributions \(N(\alpha, \sigma_\alpha^2)\), \(N(\beta_{R_m}, \sigma_{\beta_{R_m}}^2)\), \(N(\beta_{SMB}, \sigma_{\beta_{SMB}}^2)\), \(N(\beta_{BTM}, \sigma_{\beta_{BTM}}^2)\) and \(N(\beta_{MOM}, \sigma_{\beta_{MOM}}^2)\). This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different dependent variables. Total net benchmark-adjusted returns of domestic US equity investments is the dependent variable in the first row. Likewise, we regress the net benchmark-adjusted returns of large cap, small cap, actively managed and passively managed mandates on the same factors. The last column displays the results of a regression of the dependent variables on a constant, with its corresponding p-value.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Large cap</th>
<th>Small cap</th>
<th>Active</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta_{R_m}$</td>
<td>$\beta_{SMB}$</td>
<td>$\beta_{BTM}$</td>
<td>$\beta_{MOM}$</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.946)</td>
<td>(0.907)</td>
<td>(0.941)</td>
<td>(0.930)</td>
<td>(0.919)</td>
</tr>
<tr>
<td></td>
<td>-0.0011</td>
<td>-0.0035</td>
<td>0.0050</td>
<td>-0.0061</td>
<td>-0.0104</td>
</tr>
<tr>
<td></td>
<td>0.0019</td>
<td>-0.0269</td>
<td>0.0067</td>
<td>-0.0293</td>
<td>-0.0241</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.742)</td>
<td>(0.145)</td>
<td>(0.829)</td>
<td>(0.541)</td>
<td>(0.475)</td>
</tr>
<tr>
<td></td>
<td>0.0098</td>
<td>-0.0349</td>
<td>-0.1500</td>
<td>-0.0357</td>
<td>-0.0522</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.927)</td>
<td>(0.827)</td>
<td>(0.717)</td>
<td>(0.928)</td>
<td>(0.936)</td>
</tr>
<tr>
<td></td>
<td>0.0130</td>
<td>-0.0697</td>
<td>-0.0084</td>
<td>-0.1412</td>
<td>-0.0755</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.361)</td>
<td>(0.016)</td>
<td>(0.815)</td>
<td>(0.228)</td>
<td>(0.307)</td>
</tr>
<tr>
<td></td>
<td>-0.0047</td>
<td>0.0019</td>
<td>0.0029</td>
<td>0.0225</td>
<td>0.0126</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.789)</td>
<td>(0.968)</td>
<td>(0.929)</td>
<td>(0.866)</td>
<td>(0.784)</td>
</tr>
</tbody>
</table>

of defined contribution funds, which show a significant underperformance of less than 1% on an annual basis. This may be partially explained by the higher cost levels for actively managed mandates, as shown in Table 4.6. Additionally, defined contribution funds are on average smaller than defined benefit funds, i.e. they also benefit less from scale advantages in costs as explained in section 4.5.

Adding more risk factors, does not change the overall picture of performance figures. Benchmark-adjusted returns are very close to zero, before and after adding risk factors. The largely insignificant loadings on the Fama-French factors indicate that the average fund does not try to outperform its benchmark by taking on traditional risks, i.e. by investing more than the benchmark in small cap, value or momentum stocks.

Since the fund-specific benchmarks in this study are self-reported, benchmark-
4.6 Risk-adjusted Performance

Table 4.10: Performance Defined Benefit Plans

This table reports the parameter estimates of a random coefficients panel regression of net excess domestic equity returns ($R_{it,e}$) of defined benefit US pension plans on the excess S&P 500 returns ($R_{m}$), Size ("SMB"), Book-to-Market ("BTM") and Momentum ("MOM"), for the sample period 1990-2006. The random coefficients model can be characterized as

$$R_{it,e} = \alpha_i + \beta_{R_m,i} R_{m} + \beta_{SMB,i} SMB + \beta_{BTM,i} BTM + \beta_{MOM,i} MOM + \eta_{it},$$

with normally distributed mean-zero errors. We assume that the parameters $\alpha_i$, $\beta_{R_m,i}$, $\beta_{SMB,i}$, $\beta_{BTM,i}$ and $\beta_{MOM,i}$ are drawn independently from hierarchical distributions $N(\alpha, \sigma_\alpha^2)$, $N(\beta_{R_m}, \sigma_{\beta_Rm}^2)$, $N(\beta_{SMB}, \sigma_{\beta_{SMB}}^2)$, $N(\beta_{BTM}, \sigma_{\beta_{BTM}}^2)$ and $N(\beta_{MOM}, \sigma_{\beta_{MOM}}^2)$. This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different dependent variables. Total net excess return of domestic US equity investments is the dependent variable in the first row. Likewise, we regress the net excess returns of large cap, small cap, actively managed, passively managed, internally managed and externally managed mandates on the same factors.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta_{R_m}$</th>
<th>$\beta_{SMB}$</th>
<th>$\beta_{BTM}$</th>
<th>$\beta_{MOM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.0041</td>
<td>0.9358</td>
<td>0.1259</td>
<td>-0.0098</td>
<td>-0.0274</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.194)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.685)</td>
<td>(0.230)</td>
</tr>
<tr>
<td>Large cap</td>
<td>-0.0007</td>
<td>0.9497</td>
<td>0.0665</td>
<td>0.0303</td>
<td>-0.0001</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.818)</td>
<td>(0.000)</td>
<td>(0.018)</td>
<td>(0.818)</td>
<td>(0.997)</td>
</tr>
<tr>
<td>Small cap</td>
<td>0.0509</td>
<td>0.7840</td>
<td>0.4756</td>
<td>-0.3910</td>
<td>-0.3571</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Active</td>
<td>0.0064</td>
<td>0.8558</td>
<td>0.2212</td>
<td>-0.0236</td>
<td>0.0630</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.120)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.511)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Passive</td>
<td>0.0010</td>
<td>0.9253</td>
<td>0.1102</td>
<td>-0.0394</td>
<td>0.0250</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.778)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.145)</td>
<td>(0.357)</td>
</tr>
<tr>
<td>Internal</td>
<td>0.0122</td>
<td>0.8819</td>
<td>0.1402</td>
<td>0.0000</td>
<td>0.0123</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.337)</td>
<td>(0.000)</td>
<td>(0.044)</td>
<td>(1.000)</td>
<td>(0.922)</td>
</tr>
<tr>
<td>External</td>
<td>0.0053</td>
<td>0.8586</td>
<td>0.2025</td>
<td>-0.0260</td>
<td>0.0736</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.150)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.418)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>
adjusted returns may be biased as a result of strategic benchmark choices. We therefore regress net excess returns on the Fama-French factors for both defined benefit and defined contribution plans and their various mandate types. Tables 4.10 and 4.11 report the results and confirm the small but non-negative alphas documented in Tables 4.8 and 4.9. Comparing the net returns results with benchmark-adjusted performance, we can conclude that results are not affected by strategic benchmark choices. The performance picture is virtually unaffected by replacing the net benchmark return on the left-hand side net excess returns.

Ultimately, we conclude that pension funds do not consistently out- or underperform their benchmarks, but are able to cover their expenses with investment returns. This finding is in contrast with the mutual fund industry for which a large majority of studies finds negative performance after subtraction of costs (see e.g. Carhart (1997) and Malkiel (1995)).

Although our study does not suffer from biases caused by strategic benchmark choices of pension fund managers, we should note that this result may only be induced by the type of data or size of the data set. Most previous studies measure performance at the manager level using smaller cross-sections of funds. Since asset management companies may use different benchmarks than pension funds, results of studies on manager-level may not be robust to strategic benchmark choices. The benchmarking issues may also be more severe in smaller cross-sections because results are more easily affected by a small number of funds. We therefore recommend to use different market factors as a robustness test for general conclusions on performance figures in the pension industry.

4.7 Why do Pension Funds Perform Better than Mutual Funds?

The sharp contrast between pension and mutual fund alphas raises the question why pension funds perform better than mutual funds. Particularly, since Table 4.3 has shown that pension funds outsource more than 80% of their equity investments, it seems strange that pension funds have better performance figures than mutual funds. We do not argue that external managers employed by pension funds have more skill than those investing for mutual funds, since most external asset managers invest for both pension and mutual funds. We argue that scale advantages in costs for internal management, bargaining power in externally managed portfolios and better monitoring ability does not lead to outperformance of standard benchmarks, but reduces the scale and probability of underperformance in the pension sector.

Although it is difficult to test our hypothesis on the performance differential between pension and mutual funds, we could have a closer look at the gap and its constituents. The difference can to a large extent be explained by lower cost levels in the pension industry. An average pension fund has a cost level of 50 basis points, whereas mutual funds have average fees of 150 basis points. As explained in section 4.5, the lower cost level of pension funds can be partially
4.7 Why do Pension Funds Perform Better than Mutual Funds?

Table 4.11: Performance Defined Contribution Plans

This table reports the parameter estimates of a random coefficients panel regression of fund net excess domestic equity returns (\( R_{it,e} \)) of defined contribution US pension plans on excess S&P 500 return (\( R_m \)), Size (\( SMB \)), Book-to-Market (\( BTM \)) and Momentum (\( MOM \)), for the sample period 1997-2006. The random coefficients model can be characterized as

\[
R_{it,e} = \alpha_i + \beta_{Rm,i} R_m + \beta_{SMB,i} SMB + \beta_{BTM,i} BTM + \beta_{MOM,i} MOM + \eta_{it},
\]

with normally distributed mean-zero errors. We assume that the parameters \( \alpha_i, \beta_{Rm,i}, \beta_{SMB,i}, \beta_{BTM,i} \) and \( \beta_{MOM,i} \) are drawn independently from hierarchical distributions \( N(\alpha, \sigma_{\alpha}^2), N(\beta_{Rm}, \sigma_{\beta_{Rm}}^2), N(\beta_{SMB}, \sigma_{\beta_{SMB}}^2), N(\beta_{BTM}, \sigma_{\beta_{BTM}}^2) \) and \( N(\beta_{MOM}, \sigma_{\beta_{MOM}}^2) \). This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different dependent variables. Total net excess returns on domestic US equity investments are the dependent variable in the first row. Likewise, we regress the net excess returns of large cap, small cap, actively managed and passively managed mandates on the same factors.

<table>
<thead>
<tr>
<th></th>
<th>( \alpha )</th>
<th>( \beta_{Rm} )</th>
<th>( \beta_{SMB} )</th>
<th>( \beta_{BTM} )</th>
<th>( \beta_{MOM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.0114</td>
<td>0.9892</td>
<td>0.0859</td>
<td>-0.1108</td>
<td>-0.0453</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.448)</td>
<td>(0.000)</td>
<td>(0.175)</td>
<td>(0.132)</td>
<td>(0.620)</td>
</tr>
<tr>
<td>Large cap</td>
<td>0.0034</td>
<td>0.9853</td>
<td>0.0263</td>
<td>-0.0618</td>
<td>-0.0213</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.568)</td>
<td>(0.000)</td>
<td>(0.464)</td>
<td>(0.246)</td>
<td>(0.583)</td>
</tr>
<tr>
<td>Small cap</td>
<td>0.0355</td>
<td>0.9419</td>
<td>0.2608</td>
<td>-0.1260</td>
<td>-0.1586</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.740)</td>
<td>(0.000)</td>
<td>(0.525)</td>
<td>(0.754)</td>
<td>(0.808)</td>
</tr>
<tr>
<td>Active</td>
<td>0.0187</td>
<td>0.9614</td>
<td>0.1375</td>
<td>-0.1848</td>
<td>-0.0731</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.237)</td>
<td>(0.000)</td>
<td>(0.003)</td>
<td>(0.172)</td>
<td>(0.424)</td>
</tr>
<tr>
<td>Passive</td>
<td>-0.0072</td>
<td>1.0177</td>
<td>0.0194</td>
<td>0.0655</td>
<td>0.0341</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.688)</td>
<td>(0.000)</td>
<td>(0.544)</td>
<td>(0.632)</td>
<td>(0.519)</td>
</tr>
</tbody>
</table>
attributed to scale advantages in costs of internal equity investments. Pension plan participants benefit directly from these scale advantages, whereas Swensen (2005) shows that mutual fund fees are independent of fund size. Section 4.5 also provides evidence of the bargaining power of pension funds with external parties. Table 4.7 has shown that larger pension funds have relatively lower costs of external investment. This shows that larger funds are better able to bargain lower costs at asset management companies than smaller funds. Plan participants also reap the profits of the bargaining power.

In addition to costs, we argue that monitoring capacity partially explains the better performance of pension funds. Although monitoring capacity is typically hard to measure, we are able to provide evidence of the presence of monitoring abilities. When we compare the alphas of internally and externally managed portfolios in Tables 4.8 and 4.10, we find that internally managed portfolios generate higher alphas than their externally managed counterparts. Since pension funds are generally better able to monitor internal managers than external ones, this is evidence in favor of the hypothesis that better monitoring leads to better performance. Table 4.12 contains more evidence of the monitoring capacity of pension funds. The table shows parameter estimates of a panel random coefficients regression of net excess returns on log fund size and the Fama-French factors. If large pension funds are better able to monitor external parties, we expect to find scale advantages in performance, especially in the externally and actively managed mandates. In Table 4.12 we show that pension funds have positive loadings on log fund size in almost all types of equity mandates. In line with the expectations, externally and actively managed mandates show significant loadings which indicates that large funds are better able to monitor actively and externally managed mandates which in turn leads to better performance. Large pension funds often have separate accounts at external parties. Besides, they have more qualified staff to monitor the accounts. This increased monitoring ability of large funds enhances performance, as shown in Table 4.12.

4.8 Conclusion

Due to a lack of comprehensive data there exists no clear picture on the performance of US pension funds. A lack of cost and benchmark data, aggravated by severe biases in self-reported return data have lead to many different conclusions on performance. The CEM database covers a substantial part of the pension sector (40% in terms of total asset values) and contains detailed and fund-specific information on returns, costs and benchmarks. The CEM data are reported on pension plan level, contrasting the return data on manager-level in most earlier work. This fund-level data enables us to measure the performance for every individual fund. The data set also allows us to measure differences in returns and costs between large and small cap, actively and passively and externally and internally managed mandates. We rule out possible biases related to the voluntary reporting and measure pension fund performance net of fund-specific costs and benchmarks.
4.8 Conclusion

Table 4.12: Size-effect on Performance of Defined Benefit Plans

This table reports the parameter estimates of a random coefficients panel regression of net excess domestic equity returns of defined benefit US pension plans \( (R_{it,e}) \) on log fund size \( (\log(\text{Size}_{it})) \), excess S&P 500 returns \( (R_{m}) \), Size \( (SMB) \), Book-to-Market \( (\text{BTM}) \) and Momentum \( (\text{MOM}) \), for the sample period 1990-2006. The random coefficients model can be characterized as

\[
R_{it,e} = \alpha_0 + \alpha_1 \log(\text{Size}_{it}) + \beta_{R_m,i} R_m + \beta_{SMB,i} SMB + \beta_{\text{BTM},i} \text{BTM} + \beta_{\text{MOM},i} \text{MOM} + \eta_{it},
\]

with normally distributed mean-zero errors. We assume that the parameters \( \alpha_0, \alpha_1, \beta_{R_m,i}, \beta_{SMB,i}, \beta_{\text{BTM},i} \) and \( \beta_{\text{MOM},i} \) are drawn independently from hierarchical distributions \( N(\alpha_0, \sigma_{\alpha_0}^2), N(\alpha_1, \sigma_{\alpha_1}^2), N(\beta_{R_m}, \sigma_{\beta_{R_m}}^2), N(\beta_{SMB}, \sigma_{\beta_{SMB}}^2), N(\beta_{\text{BTM}}, \sigma_{\beta_{\text{BTM}}}^2) \) and \( N(\beta_{\text{MOM}}, \sigma_{\beta_{\text{MOM}}}^2) \). This table reports the parameter estimates from the hierarchical distribution and their corresponding p-values for different dependent variables. The first displays the total net excess return of domestic US equity investments of defined benefit pension plans. Likewise, we regress the net excess returns of large cap, small cap, actively managed, passively managed, internally managed and externally managed mandates on the same factors.

<table>
<thead>
<tr>
<th></th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \beta_{R_m} )</th>
<th>( \beta_{SMB} )</th>
<th>( \beta_{\text{BTM}} )</th>
<th>( \beta_{\text{MOM}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-0.1615</td>
<td>0.0230</td>
<td>0.9277</td>
<td>0.1370</td>
<td>-0.0498</td>
<td>-0.0664</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.146)</td>
<td>(0.179)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.071)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Large cap</td>
<td>-0.0989</td>
<td>0.0136</td>
<td>0.9470</td>
<td>0.0722</td>
<td>-0.0102</td>
<td>-0.0387</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.240)</td>
<td>(0.277)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.097)</td>
<td>(0.248)</td>
</tr>
<tr>
<td>Small cap</td>
<td>-0.3132</td>
<td>0.0550</td>
<td>0.7506</td>
<td>0.4562</td>
<td>-0.4250</td>
<td>-0.3832</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.309)</td>
<td>(0.147)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Active</td>
<td>-0.3738</td>
<td>0.0482</td>
<td>0.8505</td>
<td>0.2257</td>
<td>-0.0922</td>
<td>-0.0005</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.020)</td>
<td>(0.991)</td>
</tr>
<tr>
<td>Passive</td>
<td>-0.1927</td>
<td>0.0226</td>
<td>0.9279</td>
<td>0.1195</td>
<td>-0.0567</td>
<td>0.0237</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.086)</td>
<td>(0.116)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>(0.130)</td>
<td>(0.393)</td>
</tr>
<tr>
<td>External</td>
<td>-0.4113</td>
<td>0.0527</td>
<td>0.8444</td>
<td>0.2286</td>
<td>-0.0999</td>
<td>0.0257</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.498)</td>
</tr>
<tr>
<td>Internal</td>
<td>0.0059</td>
<td>-0.0013</td>
<td>0.8764</td>
<td>0.1524</td>
<td>-0.0189</td>
<td>0.0351</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.987)</td>
<td>(0.975)</td>
<td>(0.000)</td>
<td>(0.040)</td>
<td>(0.837)</td>
<td>(0.774)</td>
</tr>
</tbody>
</table>
We find that pension fund net returns are very close to self-reported benchmarks. Cost levels are substantially lower than standard mutual fund fees and are diminishing in fund size. The lower cost levels can be partially explained by a combination of scale advantages in internally managed portfolios and bargaining power of large pension funds with external management companies. Plan participants benefit directly from the lower costs, whereas mutual fund investors do not harvest these profits as a result of the mutual fund profit objective. We also show evidence of the monitoring power of large pension funds, which improves performance, again to the benefit of pension plan participants. Additionally, we give an overview of the benchmarks used by pension funds to evaluate their asset managers. We show that pension funds use many different benchmarks both across mandates and over time. By using different market factors in our performance evaluation, we point out that the effects of choosing the wrong benchmark can be substantial. We consequently contribute to earlier work by showing that our results are robust to choices of different benchmark, whereas previous studies only benchmark performance against the S&P 500.
Chapter 5

Regret Aversion and Annuity Risk in Defined Contribution Pension Plans

The high value of the implicit option to choose a retirement date at which interest rates are particularly high and life annuities relatively cheap, leads to the possibility to introduce regret aversion in the retirement investment decision of defined contribution plan participants. As a remedy for regret aversion in retirement investment decisions, this chapter develops and prices a lookback option on a life annuity contract. We determine a closed-form option value under the restriction that the option holder invests risklessly during the time to maturity of the option and without the guarantee that the exact amount of retirement wealth is converted into a life annuity at retirement. Thereafter the investment restriction is relaxed and the guarantee of exact conversion is imposed and the option is priced via Monte Carlo simulations in an economic environment with a stochastic discount factor. Option price sensitivities are determined via the pricing of alternative options. We find that the price of a lookback option, with a maturity of three years, amounts to 8-9% of the wealth at the option issuance date. The option price is highly sensitive to the exercise price of the option, i.e. pricing alternative options (e.g. Asian) substantially lowers the price. Time to maturity and interest rate volatility are other important option price drivers. Asset allocation decisions and initial interest rates hardly affect the option price.\(^1\)

5.1 Introduction

In recent years an increasing number of pension funds shifted from defined benefit to defined contribution plans. Particularly in anglo-saxon countries defined

\(^1\) This chapter is based on the paper Frehen, Hoevenaars, Palm, and Schotman (2008)
contribution plans have gained popularity. Boulier, Huang, and Taillard (2001) attribute this shift to two major advantages of defined contribution plans over traditional defined benefit plans. Participants can observe their pension claim at any moment in time and can transfer this claim more easily when changing employer. Moreover, defined contribution plan sponsors are unburdened from risk associated with the pension plan. This risk is passed on to participants. As a direct result of the change in risk bearer, large numbers of uninformed participants face a difficult decision at retirement: Whether and when should they invest their retirement wealth irreversibly in a life annuity? The high price sensitivity of the life annuity with respect to the long-term interest rate complicates the retirement investment decision. In times of increasing interest rates, plan participants may be forced to postpone their retirement date. Beside the shorter payout period and longer contribution period, the high price sensitivity of the life annuity can offer substantial gains in times of increasing interest rates. Vice versa, plan participants may also decide to retire early in times of high interest rates. Hence, the high price sensitivity of the life annuity makes the implicit option to time the retirement date valuable (see Milevsky and Young, 2006).

The large value of the timing option creates fertile ground for regret aversion. For instance, the pension income of a participant who retired at the end of our data set (December 2003) was, ceteris paribus, 6.1% higher than a colleague’s income who retired half a year earlier. The 6.1% difference is even more remarkable if we take into account that long-term interest rate differences at the sample end are small relative to differences earlier in the sample (see Figure 5.1).

Numerous experiments document that regret aversion plays a role in human decision making (e.g. Kahneman and Tversky, 1979). Bell (1982) and Loomes and Sugden (1982) are the first to extend standard utility functions by incorporating regret aversion components. Since this development, the effect of regret aversion has been explored in many different areas. Braun and Muermann (2004) document that regret aversion has a mitigating effect on extreme demands for insurance. Michenaud and Solnik (2008) show that even if the currency risk premium is zero, regret averse investors hold positive currency positions. Gollier and Salanié (2006) document that the introduction of regret aversion in a complete market with Arrow-Debreu securities, shifts the optimal asset allocation more to states with low probabilities. Muermann, Mitchell, and Volkman (2006) analyze the role of regret aversion in pension investment decisions. They evaluate the effect of regret aversion on the willingness to pay for a pension guarantee. One of their main findings is that the willingness to pay for a pension guarantee, of an investor with a risky portfolio, increases with the introduction of regret aversion.

This chapter prices an option that protects regret averse participants against annuity risk at retirement. At retirement, many retirees convert their pension wealth into a life annuity and regret aversion plays an important role in the timing of the life annuity purchase. Merton and Bodie (2005) argue that lookback options are valuable assets for regret averse individuals. We address regret aversion and annuity risk at retirement by developing and pricing a lookback option on a life annuity. The option holder buys an insurance against regret aversion some
5.1 Introduction

Figure 5.1: Conversion Rate

This figure displays the time series evolution of the 10 year zero coupon bond yield from January 1952 to January 2004. The yields are annualized.
years before the retirement date. Using a replicating portfolio strategy, we derive the closed-form price for the lookback life annuity option under some simplifying assumptions. We also determine the option price sensitivity with respect to the volatility of the underlying interest rate process. In the closed-form setting, the option holder is obliged to invest in a riskless product during the time to maturity of the option. Furthermore, the closed-form solution cannot guarantee that the exact amount of retirement wealth is converted into a life annuity for the minimum forward price. As a result, the closed-form solution does not eliminate annuity risk completely.

To allow risky investment strategies and eliminate all annuity risk, we also price the lookback life annuity option in an extended economic environment. In this more elaborate setting, we price the option under the restriction that the complete retirement wealth is converted into a life annuity and allow stock and bond investments during the time to maturity of the option. Monte Carlo simulations and a stochastic discount factor are used to price the option. Finally, we determine the option Greeks by changing different parameters in the pricing environment.

We find that the closed-form price of the lookback life annuity option amounts to 7-9% of the wealth at the option issuance date and that interest rate volatility and time to maturity are important price drivers. A relaxation of some simplifying assumptions and the application of a stochastic discount factor results in a comparable option price of 8% for men and 9% for women. This relatively high price indicates that the insurance against regret aversion is expensive. It is therefore unlikely that optimization of life cycle functions ignoring regret aversion leads to life annuity lookback option purchases. Hence accounting for regret aversion results in fundamentally different investment strategies than observed in standard non-regret averse investor strategies. Furthermore, we document that the lookback feature is an important driver of the option price. If the participant is offered the right to buy the life annuity for the average forward life annuity price (Asian option), the option price decreases to slightly more than 2% of the wealth at the option issuance date. Extending the time to maturity of the option by one year increases the option price to 11-12%. We also find that the option price is insensitive to the initial interest rate and hardly affected by the asset allocation during the time to maturity.

The presence of regret aversion in retirement investment decisions contrasts this chapter sharply with most earlier literature on the topic. The lookback life annuity option, which offers protection against regret aversion, is not available in financial markets yet. Earlier literature addresses the problem of annuity risk at retirement effectively using standard utility functions and asset allocation approaches. Assuming power utility for participants, Campbell and Viceira (2002) derive the optimal pre-retirement asset allocation. Furthermore, Yaari (1965) points out that it is optimal for participants without bequest motives to convert retirement wealth into a life annuity. Koijen, Nijman, and Werker (2008) determine the optimal allocation to nominal, real and equity-linked annuities at retirement. Subsequently, they determine the optimal pre-retirement hedging strategy in stocks, nominal and inflation-linked bonds and cash, for the risk created by the optimal annu-

104
5.2 Defined Contribution and the Lookback Option

In this section we develop and price the lookback option on a life annuity contract in a continuous time framework. The lookback option is offered to plan participants three years before the retirement age of 65. We assume that participants who buy the option survive until their retirement date. Moreover, we price the option under the assumption of exogenous and constant mortality rates and fairly priced life annuities. As a result, a life annuity is nothing more than a portfolio of zero-coupon bonds. The pricing of the option can therefore be reduced to pricing a series of lookback options and each is on a pure discount bond. We therefore start with the pricing of a lookback option on a zero coupon bond with fixed maturity date \( s \). We consider a fixed option issuance date \( t_a \). This allows us to suppress the dependence of mortality rates and realized forward bond prices on \( t_a \). Bond option formulas developed by Jamshidian (1989) and Goldman, Sosin, and Gatto (1979) are used as a guideline to price the lookback option. We assume that the instantaneous spot rate follows an Ornstein-Uhlenbeck interest rate process characterized as

\[
dr = \alpha (\gamma - r) dt + \sigma_r dW,
\]

with \( r \) being the instantaneous spot rate, \( \alpha \) the mean-reversion parameter, \( \gamma \) the long-term average interest rate, \( \sigma_r \) the interest rate volatility and \( dW \) a standard Brownian motion under the risk neutral measure.
To characterize the lookback option on a discount bond we specify the option payoff. Since we price the option via a replicating portfolio technique this characterization is sufficient to determine a closed-form option price. The lookback bond option provides the right at the fixed retirement date $T$ to buy a discount bond with maturity date $s$ for the minimum forward price. The lookback bond option payoff $U_T$ can consequently be characterized as

$$U_T = P_rT, T, s) - Q(f_T, T, s),$$  \hfill (5.2)

with $P_rT, T, s)$ the time $T$ price of a bond maturing at $s$ and $Q(f_T, T, s)$ the minimum forward bond price in the time span $[t_a, T]$, where the dependence of $Q$ on $t_a$ is suppressed. The minimum forward bond price is a function of the maximum forward rate $f_T$ in the time interval $[t_a, T]$, retirement date $T$, and bond maturity date $s$.

An investor who wants to replicate the lookback option on an $s$-maturity discount bond must buy a straddle on the $s$-maturity bond with exercise price equal to the initial (time $t_a$) forward bond price at time $T$. Once the time $T$ forward bond price attains a new minimum, the replicating investor sells the old position and buys a new straddle with the new minimum as exercise price. He repeats this exercise until the straddle matures. The time $t$ exercise price $Q(f_t, t, s) = Q$ of the straddle is consequently the minimum forward bond price up till time $t$. $Q$ thus summarizes the history of the lookback option and therefore contains the lookback-element. The bond prices and therefore also the bond option prices are direct contingents of the time $t$ interest rate. Moreover, the option and bond maturity dates $T$ and $s$ are fixed. The hedging portfolio value can consequently be characterized as a function of the time $t$ interest rate $r$, exercise price $Q$, and time $t$, and follows

$$H(r, Q, t) = \text{Call}_B + \text{Put}_B,$$  \hfill (5.3)

with $\text{Call}_B$ and $\text{Put}_B$ respectively the time $t$ prices of bond call and put options maturing at time $T$ on pure discount bonds with maturity $s$. A simple application of the put-call-parity leads to

$$H(r, Q, t) = 2\text{Call}_B + P_rT, T, T)Q - P_rT, t, T).$$  \hfill (5.4)

Substitution of bond option formulas by Jamshidian (1989) shows that the bond straddle that replicates the lookback option is composed of investments in discount bonds with maturities $T$ and $s$. This is a result of the fact that standard put and call bond options can be replicated with investments in bonds maturing at times $T$ and $s$. The hedging portfolio for the lookback option on an $s$-maturity bond can therefore be simplified to two bond investments and can be expressed as

$$H(r, t, Q) = \phi_t^{(s)} P_rT, t, s) + \phi_t^{(T)} P_rT, t, T),$$  \hfill (5.5)

with

$$\phi_t^{(T)} = [1 - 2N(h - \sigma_p)]Q,$$  \hfill (5.6)

$$\phi_t^{(s)} = 2N(h) - 1,$$  \hfill (5.7)
5.2 Defined Contribution and the Lookback Option

where \( N \) represents the normal cumulative distribution function and Jamshidian (1989) characterizes \( h \) and \( \sigma_p \) as

\[
\begin{align*}
    h &= \ln \left( \frac{P(r,t,s)}{P(r,t,T)} \right) + 1 + \frac{\sigma_p}{2} = \ln \left( \frac{P(r,t,s)}{P(r,t,T)} \right) + \frac{\sigma_p^2}{2} + 1, \\
    \sigma_p &= v(t,T) \left( 1 - e^{-\alpha(s-T)} \right), \\
    v^2(t,T) &= \sigma^2 1 - e^{-2\alpha(T-t)}.
\end{align*}
\]

(5.8) (5.9) (5.10)

In the appendix we prove that the hedging portfolio is indeed a replicating strategy for the lookback bond option. Hence, the price of the hedging portfolio must equal the price of the lookback bond option. The time \( t_a \) price of the lookback option on a zero coupon bond maturing at time \( s \) can therefore be characterized as

\[
C^B_{L}(P(r,t_a,s),Q,T-t_a,s) = \phi^{(s)}_{t_a} P(r,t_a,s) + \phi^{(T)}_{t_a} P(r,t_a,T).
\]

(5.11)

Under certain conditions Jamshidian (1989) proves that the value of a portfolio of bond options equals the value of an option on a bond portfolio. Consequently, the price of a lookback call option on a life annuity is defined as

\[
C^L_{L}(r,\tilde{\phi},t_a,T) = \int_{t_a}^{T} C^B_{L}(P(r,t,s),Q,T-t_a,s) \zeta_s ds,
\]

(5.12)

with \( \zeta_s \) the time \( t \) conditional survival density, i.e. if the time \( t \) probability that the participant is still alive at time \( s \) is modeled as \( \pi_s \), then its derivative with respect to \( s \) is denoted \( \zeta_s = \frac{\partial \pi_s}{\partial s} \).

This closed-form option pricing formula enables us to price a lookback option on a life annuity with a monthly nominal payoff equal to 1. Table 5.1 displays the percentage of wealth at the option issuance date \( (t_a) \) that a participant pays for the right to buy a yearly pension income of 10,000 for the minimum price. The numbers in Table 5.1 are obtained by discretizing equation (5.12) on a monthly basis and estimating equation (5.1) with monthly US one-month interest rate data from January 1952 to December 2003 and substituting the estimates for the parameters in equation (5.12).

Table 5.1 shows that time to maturity is an important driver of the lookback life annuity option price. Furthermore the table shows that for a time to maturity of three years a plan participant pays 7-9% of the wealth at the option issuance date for the right to convert his retirement wealth against the minimum forward life annuity price during the time to maturity of the option.

A closed-form solution also enables us to estimate the effect of a change in standard deviation of the underlying interest rate model on the option price. Figure 5.2 displays option prices for volatilities of the interest rate process ranging from 50% to 150% of the estimated volatility.
Figure 5.2: Changes in Interest Rate Volatility

This figure displays the effect of a change in interest rate volatility on the option price. Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women. Prices are provided for volatilities ranging from 50% to 150% of the estimated volatility.
5.3 Extended Environment and Price Drivers

Table 5.1: Base Closed-Form Option Prices

This table presents prices for the lookback life annuity option with times to maturity of respectively two, three and four years and a yearly payoff of 10,000. Prices for both men and women are expressed in percentages of the wealth at the issuance date of the option \( W_{\text{t}} \).

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>5.93</td>
<td>7.64</td>
<td>8.90</td>
</tr>
<tr>
<td>Women</td>
<td>7.59</td>
<td>9.60</td>
<td>11.10</td>
</tr>
</tbody>
</table>

Figure 5.2 shows that the volatility of the instantaneous spot rate is an important driver of the option price. Decreasing the volatility by 50% leads to a decrease in option price of approximately 50%. Vice versa, increases in the volatility of the spot rate of 50% lead to option price increases of approximately 50%.

Figure 5.3 displays the effects of the changes in hedging portfolio holdings, due to a rise in interest rate from 2% to 11%. We zoom in on the time \( t \) holdings of a three \( (\phi_{t}^{(T)}) \) and four \( (\phi_{t}^{(s)}) \) year maturity bond, assuming a fixed minimum interest rate of 2% and fixed time to maturity of three years.

Figure 5.3 shows that, ceteris paribus, a rise in interest rate leads to larger long positions in the bond with three year maturity and larger short positions in a four year maturity bond. The figure also shows an upper- and lowerbound on bond holdings. These upper- and lowerbounds can easily be explained by equations (5.6) and (5.7). The cumulative normal distribution implies a range of \([-Q, Q]\) for \( \phi_{T} \) and a range of \([-1, 1]\) for \( \phi_{s} \), as clearly depicted in Figure 5.3.

One limitation of the closed form model is that it works with an option payoff that is per unit of annuity income after retirement. The participant buys a fixed number of options, and each option provides the payoff (5.2). When the participant invests his wealth in risky assets, the final wealth at age 65 is unknown, and therefore he does not know a priori how many option he needs to convert his entire pension wealth at the retirement date. An alternative design of the option is that it gives the right to convert all wealth in the pension account at the retirement.

5.3 Extended Environment and Price Drivers

In this section we describe the pricing of a lookback life annuity option that guarantees conversion of final wealth in a pension account. The pricing environment generalizes the one-factor Ornstein-Uhlenbeck assumption from the previous section to a multi-factor setting. We also allow the participant to invest part of his pension wealth in stocks. Closed-form solutions are no longer available in this extended setting and we therefore price the option using Monte Carlo methods and a stochastic discount factor that is consistent with market prices for stocks and bonds. If the participant can choose his investment portfolio (stocks, bonds...
Figure 5.3: Changes in Hedging Portfolio Holdings

This figure displays the effect of a change in three and four year bond holdings due to a rise in interest rate from 2\% to 11\%. We consider a fixed time to maturity of three years and fixed minimum interest rate of 2\% Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women.
of different maturities), he also controls the expected value and volatility of his final pension wealth. This in turn implies that the price of the option depends on the investment strategy chosen by the participant. As we work with simulations, we also switch from continuous to discrete time.

5.3.1 Pricing Environment

This subsection describes the economic environment and pricing techniques applied to determine the lookback life annuity price. We price the option using Monte Carlo methods, again assuming that mortality rates are exogenous and constant, fair pricing of life annuities and neglecting mortality risk and contributions during the time to maturity of the option (as in the previous section). The probability that a participant dies during the time to maturity of the option is approximately 3%. Including mortality risk during the time to maturity of the option, would therefore decrease the option price by approximately 3%. We neglect the effect of the lookback option purchase on the pre-retirement wealth. Despite its implausible character, this assumption is necessary. Otherwise, the retirement wealth and thus the option price would be a function of the option price. This self-dependency is typically hard to model. In the simulation setting the retirement wealth \( W_T \) is a function of the investment weights, specified by the option holder and the accompanying future returns which are uncertain at time \( t_a \),

\[
W_T = A_{t_a} \prod_{t=t_a}^{T} (w_s (1 + R_{s,t}) + w_{2b} (1 + R_{2b,t}) + w_{10b} (1 + R_{10b,t})), \tag{5.13}
\]

with \( w_s \), \( w_{2b} \) and \( w_{10b} \) respectively the investment weights for the stock portfolio, two-year and ten-year bond and \( R_{s,t} \), \( R_{2b,t} \) and \( R_{10b,t} \) the corresponding time \( t \) returns. \( A_{t_a} \) represents the participant’s initial wealth. In the base simulation setting, we set \( A_{t_a} \) equal to 100,000 and specify equal investment weights \( (w_s = w_{2b} = w_{10b}) \), whereas the returns are simulated using a VAR-model. We first determine the option payoff and specify the processes for the stock and bond returns in the investment universe with the corresponding stochastic discount factor. Subsequently, 100,000 scenarios with discounted option payoffs are generated on a monthly frequency. The option price is then approximated by the average discounted option payoff.

To ultimately determine the option price, the option payoff, investment universe and discount factor need to be specified. In this setting option holders have the right to buy a life annuity for the minimum forward price attained during the time to maturity of the option. Furthermore, we require that both the option holder and non-option holder convert the exact amount of retirement wealth \( W_T \) into a life annuity. As a result, the right to buy the life annuity for minimum price is equivalent to the right to convert retirement wealth against the maximum interest rate during time to maturity of the option. We specify the guarantee of
exact conversion of retirement wealth for non-option holders as

\[ W_T = L^N \sum_{s=T+1}^{\infty} \frac{\pi_s}{(1 + r_T)^{s-T}}, \tag{5.14} \]

where \( \pi_s \) is the time \( T \) probability that the participant is still alive at time \( s \) and \( L^N \) the after retirement income of the non-option holder (i.e. the nominal monthly life annuity payoff for the non-option holder). The option holder has the right to convert against the maximum rate. The guarantee for the option holder is therefore expressed as

\[ W_T = L^H \sum_{s=T+1}^{\infty} \frac{\pi_s}{(1 + r_{\text{max}})^{s-T}}, \tag{5.15} \]

where \( L^H \) is the after retirement income of the option holder (i.e. the nominal monthly life annuity payoff of the option holder) and \( r_{\text{max}} \) the maximum interest rate during the time to maturity of the option. The option payoff is the difference between the present value of the option holder’s and non-option holder’s benefits

\[ U_T = \sum_{s=T+1}^{\infty} \frac{\pi_s [L^H - L^N]}{(1 + r_T)^{s-T}} \]

Equations (5.15) and (5.16) show that the option payoff is a function of the retirement conversion rate \( r_T \), maximum conversion rate \( r_{\text{max}} \), terminal wealth \( W_T \) and mortality rates \( \pi_s \).

In a defined contribution pension plan, participants have more freedom with regard to the asset allocation than in a defined benefit plan. Defined contribution plan participants have to specify weights \( (w_s, w_{2b} \text{ and } w_{10b}) \) for each product (portfolio of stocks, two-year bond and ten-year bond) in the investment universe. The investment universe with stocks and different bonds implies that investment risk during the time to maturity of the option is contained in the option price, in contrast to the closed-form analysis. Analogous to Ang and Piazzesi (2003) and Cochrane and Piazzesi (2005) we model the log returns of the different investment alternatives as a VAR(1)

\[ X_{t+1} = \mu + \phi X_t + \eta_{t+1}, \]

As Campbell, Lo, and MacKinlay (1997), we approximate \( n \)-maturity log bond returns by

\[ r_{n,t+1} = D_{n,t}y_{n,t} - (D_{n,t} - 1)y_{n-1,t+1}, \tag{5.18} \]
5.3 Extended Environment and Price Drivers

Table 5.2: Return Summary Statistics

This table presents summary statistics of the annualized stock, two and ten-year bond returns. For each return series, different rows display the mean, standard deviation, skewness and kurtosis. The statistics are computed on a data period ranging from January 1952 until January 2004.

<table>
<thead>
<tr>
<th></th>
<th>$r_s$</th>
<th>$r_{2b}$</th>
<th>$r_{10b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.108</td>
<td>0.059</td>
<td>0.065</td>
</tr>
<tr>
<td>St. Dev</td>
<td>0.149</td>
<td>0.100</td>
<td>0.093</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.708</td>
<td>0.840</td>
<td>0.713</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.79</td>
<td>0.762</td>
<td>0.188</td>
</tr>
</tbody>
</table>

with $D_{n,t}$ and $y_{n,t}$ respectively the time $t$ duration and yield of an $n$-maturity bond. We approximate $y_{n-1,t+1}$ by $y_{n,t+1}$ and consider zero coupon bonds, for which maturity equals duration.

The VAR(1) model is estimated using monthly U.S. data on bond yields and stock returns for the period ranging from January 1952 to December 2003. Value weighted stock returns are obtained from the CRSP database. Bond yields are provided by Bliss (1997). Summary statistics and parameter estimates are reported in Tables 5.2 and 5.3.

In order to price the option, we specify a stochastic discount factor. The stochastic nature of the discount factor allows it to discount payoffs in relatively “good” scenarios stronger than payoffs in relatively “bad” scenarios. In line with Cochrane and Piazzesi (2005) we assume for the discount factor the following process

$$M_{t+1} = \exp \left(-\delta_0 - \delta_1' X_t - \frac{1}{2} \lambda_t' \Sigma \lambda_t - \lambda_t' \eta_{t+1} \right),$$  \hspace{1cm} (5.19)

$$\lambda_t = \lambda_0 + \lambda_1' X_t,$$ \hspace{1cm} (5.20)

where $M_{t+1}$ denotes the time $t+1$ discount factor and $\lambda_t > 0$. By definition, the expected deflator value should equal the price of a zero coupon bond maturing in the same period. Considering a one period horizon, this yields

$$p_t^{(1)} = \log \left( E_t \left[ M_{t+1} \right] \right),$$ \hspace{1cm} (5.21)

with $p_t^{(1)}$ the log time $t$ price of a zero coupon bond with maturity date $t + 1$. Substitution of equations (5.19) and (5.20) in equation (5.21) leads to

$$-\delta_0 - \delta_1' X_t = -y_{1m},$$ \hspace{1cm} (5.22)

with $y_{1m}$ the log yield of a one month maturity bond at time $t$. To determine values for $\lambda_0$ and $\lambda_1$, we note that the expected value of the discounted return of any asset in the economic environment should be equal to 1, represented as

$$\mu = E_t \left[ M_{t+1} R_{t+1} \right],$$ \hspace{1cm} (5.23)
Table 5.3: VAR Estimation Results

This table presents parameter estimates and the corresponding t-statistics of the VAR(1)-model for stock and bond returns and for $\delta_0$ and $\delta_1$ as specified in section 5.3. Log monthly stock, two-year and ten-year bond returns are respectively denoted by $r_{s,t}$, $r_{2b,t}$ and $r_{10b,t}$ and $c$ represents a constant. The one month log bond yield is represented by $y_{(1)1mt}^{(1)}$. We approximate the one month maturity bond yield by the monthly yield on a three month maturity bond, for data reasons. t-statistics are reported in parentheses. The estimations are performed on a data period ranging from January 1952 until January 2004.

<table>
<thead>
<tr>
<th></th>
<th>$c$</th>
<th>$r_{s,t-1}$</th>
<th>$r_{2b,t-1}$</th>
<th>$r_{10b,t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{s,t}$</td>
<td>0.0055</td>
<td>0.0490</td>
<td>0.5806</td>
<td>0.0705</td>
</tr>
<tr>
<td></td>
<td>(2.5604)</td>
<td>(1.2202)</td>
<td>(1.6162)</td>
<td>(0.4917)</td>
</tr>
<tr>
<td>$r_{2b,t}$</td>
<td>0.0040</td>
<td>-0.0330</td>
<td>0.2101</td>
<td>0.0116</td>
</tr>
<tr>
<td></td>
<td>(10.7840)</td>
<td>(-4.7172)</td>
<td>(3.3589)</td>
<td>(0.4628)</td>
</tr>
<tr>
<td>$r_{10b,t}$</td>
<td>0.0037</td>
<td>-0.0653</td>
<td>0.4208</td>
<td>-0.0485</td>
</tr>
<tr>
<td></td>
<td>(3.8897)</td>
<td>(-3.6423)</td>
<td>(2.6217)</td>
<td>(-0.7573)</td>
</tr>
<tr>
<td>$y_{(1)1mt}^{(1)}$</td>
<td>0.0036</td>
<td>-0.0044</td>
<td>0.1912</td>
<td>-0.0455</td>
</tr>
<tr>
<td></td>
<td>(35.0430)</td>
<td>(-2.2660)</td>
<td>(11.0330)</td>
<td>(-6.5832)</td>
</tr>
</tbody>
</table>

with $\iota$ a 3x1 unit vector and $R_{t+1}$ a 3x1 vector containing the returns of the different products in the investment universe. Rewriting and taking logs allows us to rewrite equation (5.23) as

$$0 = \iota E_t \{ \log (M_{t+1}) \} + E_t \{ \log (R_{t+1}) \] + \frac{1}{2} \sigma_m^2 + \frac{1}{2} \text{diag} (\Sigma) + \sigma_{mr},$$

(5.24)

with $\sigma_m^2$ the variance of the log discount factor, $\text{diag}(\Sigma)$ a $3 \times 1$ vector containing the diagonal elements of $\Sigma$ and $\sigma_{mr}$ a $3 \times 1$ vector containing the covariances between the log discount factor and the log returns of respectively the stock, 2 year bond and 10 year bond. Substitution and some more rewriting then gives

$$\lambda_t = \Sigma^{-1} \left( \mu + \phi X_t + \iota \left( -\delta_0 - \delta_1' X_t \right) + \frac{1}{2} \text{diag} (\Sigma) \right),$$

(5.25)

$\lambda_0$ and $\lambda_1$ can consequently be defined as

$$\lambda_0 = \Sigma^{-1} \left( \mu - \delta_0 \iota + \frac{1}{2} \text{diag} (\Sigma) \right),$$

(5.26)

$$\lambda_1 = \Sigma^{-1} \left( \phi - \iota \delta_1' \right),$$

(5.27)

where values for $\delta_0$ and $\delta_1$ have been obtained by simply regressing the log short-term interest rate yield on the VAR-variables $X_t$ and $\mu$ and $\phi$ and $\Sigma$ from the VAR-regression.
5.3 Extended Environment and Price Drivers

We have specified the option payoff in equation (5.16) and discount factor in equation (5.19) in VAR parameters, mortality rates and terminal wealth. VAR parameters have been obtained from standard regressions and mortality rates are readily available. However terminal wealth depends on the investment weights and corresponding returns in each investment category for the investor. Hence, by specifying weights \( w_s, w_{2b} \) and \( w_{10b} \) for each investment category and repetitively, randomly drawing values for \( \eta \) from a normal distribution, scenarios-specific values for the terminal wealth and thus for the option payoff and discount factor can be determined. The option price can subsequently be approximated by averaging the discounted option payoffs across all scenarios.

Any \( n \)-maturity bond can be priced consistently with the VAR specified before. This allows us to determine the exact life annuity price and easily extend the investment universe. The exact price is obtained by discounting each pension payoff with the corresponding maturity bond rate. The investment universe as specified contains two bonds, a two-year and ten-year maturity zero coupon bond. This may seem a restrictive universe. However, Cochrane and Piazzesi (2005) document that a complete term structure of bonds can be defined recursively. As a result, it is straightforward to extend the universe with bonds of any maturity. However, we have to note that the term structure in our investment universe is based on only two maturities. As a result, the empirical term structure could differ (even in shape) from ours. If we assume that the \( n \)-maturity bond yield is a linear function of the VAR-variables

\[
y_t^{(n)} = a_n + b'_n X_t,
\]

and use the fact that \( A_0=0 \) and \( B_0=0 \), \( A_n \) and \( B_n \) can be defined recursively using

\[
\begin{align*}
A_{n+1} &= -\delta_0 + A_n + B'_n \mu - B'_n \Sigma \lambda_0 + \frac{1}{2} B'_n \Sigma B_n, \\
B'_{n+1} &= -\delta'_1 + B'_n \phi - B'_n \Sigma \lambda_1.
\end{align*}
\] (5.29)

In line with Cochrane and Piazzesi (2005) \( a_n \) and \( b'_n \) are obtained as

\[
\begin{align*}
a_{n+1} &= -\frac{1}{n+1} \left( -\delta_0 + A_n + B'_n \mu - B'_n \Sigma \lambda_0 + \frac{1}{2} B'_n \Sigma B_n \right), \\
b'_{n+1} &= -\frac{1}{n+1} \left( -\delta'_1 + B'_n \phi - B'_n \Sigma \lambda_1 \right).
\end{align*}
\] (5.31)

5.3.2 Measuring Price Drivers with Alternative Options

This subsection describes the different alternative options priced to quantify the Greeks of the lookback life annuity option. We start with a base option and first modify the exercise price (i.e. the option version) to determine the effect of the lookback characteristic. Subsequently the parameters time to maturity, initial interest rate and investment weights are changed to resolve the price sensitivities of the option price.
5 Regret Aversion in Defined Contribution Pension Plans

Base Option

The base option is a lookback life annuity option with a time to maturity of three years. This option provides plan participants the right to buy a life annuity at retirement for the minimum forward price during the last three years before retirement. In addition to the time horizon, the option price is a function of the retirement wealth $W_T$. $W_T$ is determined by the investment weights ($w_s$, $w_{2b}$ and $w_{10b}$) and their corresponding returns. We assume that contributions do not take place in the last years before retirement and assume that participants that buy the option survive until retirement. For the base option we assume that investment weights are equal ($w_s = w_{2b} = w_{10b}$) and that the portfolio is monthly rebalanced.

As a final characterization, we set the initial value of the conversion rate equal to its last value in the dataset (4.42%).

Option Greeks

As a first sensitivity test, we change the exercise price to determine the impact of the lookback feature on the price. We price an Asian option with the same characteristics as the lookback option. The Asian option provides the right to buy the life annuity for the average forward price during the time to maturity. This option only has a positive payoff if the time $T$ life annuity price exceeds the average forward life annuity price. The second alternative option we price has a predetermined and fixed exercise price. This option offers holders the right to convert their retirement wealth against a predetermined rate (e.g. 6%). The option holder exercises the option if the retirement rate does not exceed the guaranteed rate. For both alternatives the probability of having a payoff of zero is substantial, whereas the lookback option has a negligible chance of ending at-the-money (and generating a zero payoff). Since the option price is the expected value of the discounted payoff, the effect of changing the option version is large.

Furthermore, we measure the effect of a change in time to maturity. We price the lookback life annuity option on a horizon of two, four and five years. By providing the right to buy for the minimum price, the time to maturity is an important driver of the option price. An extension in time horizon with one year, gives the life annuity price 12 additional possibilities of attaining a new minimum. Hence, the longer (shorter) the time to maturity, the larger (smaller) the expected option payoff and thus the option price.

As a further possible price driver we select the interest rate at the time the option is issued. Particularly if the interest rate three years before retirement is high, a large demand for the lookback option is likely to arise. Participants then want to protect themselves against a fall in interest rate shortly before the retirement date. We quantify the impact of a change in initial rate. However, we point out in the closed-form valuation (see appendix) that changes in the lookback option price are not affected by changes in minimum price of the life annuity. On the issuance date $t_a$, the life annuity price is at a minimum by construction. Hence, the option price is unlikely to be affected by changes in the initial rate. The guarantee that the exact amount of retirement wealth is converted does not
Finally, we measure the impact of the investment weights on the option price. In addition to the equally weighted portfolio of the base option, we construct three pure asset portfolios. Investments are restricted to one of the three assets (stock portfolio, two-year and ten-year bond) in the pure portfolios. The portfolios are re-balanced on a monthly basis. Since we consider the ten-year bond yield as the conversion rate of the life annuity, especially the pure ten-year bond portfolio is an interesting investment strategy. This portfolio is particularly risky. When the interest rate increases during the time to maturity of the option, the difference in interest rates \((r_T - r_{max})\) is likely to be small and the option payoff consequently low. Additionally, interest rate increases lead to a relatively low retirement wealth \(W_T\), via low bond returns. On the other hand, interest rate decreases lead to large differences between \(r_T\) and \(r_{max}\) and thus high option payoffs. The high payoff is amplified by a relatively large retirement wealth. The large retirement wealth is a result of high bond returns. The higher risk of the pure ten-year bond investment strategy is partially offset by the stochastic discount factor. Nevertheless, investing in the pure ten-year bond portfolio is a risky strategy.

5.4 Empirical Results

This section reports the prices of the previously specified options. We start with pricing a base option and then provide prices for different alternatives, to determine the option Greeks.

5.4.1 Option Prices

The base option as described in the previous section is a lookback life annuity option with a time to maturity of three years. Since men and women have different mortality rates, we determine option prices for men and women separately.

Table 5.4 reveals that the price of the base option is 8% for male participants and slightly higher than 9% for female participants, due to better life expectancies for women. Furthermore, the simulation results show that the standard deviation across all scenarios of the discounted option payoff is high, approximately 55% for men and 63% for women. This is an indication that the annuity risk that participants encounter at the conversion date is considerable. Compared to the results in section 5.2, we document a price increase of 2-3% points. The increase can be attributed to the risky investment strategy and the guarantee that the exact amount of retirement wealth is converted at retirement. Table 5.4 also shows convergence in option prices at 100,000 simulations. Furthermore, 100,000 simulations ensure significance up to the second decimal. We therefore hereafter generate prices based on 100,000 simulations. Finally, the table with base option prices displays the average across all scenarios of the deflated cumulative stock returns. As expressed by equation (5.23) the expected value of discounted cumulative stock returns should be equal to one. We approximate the expected value by the average across all scenarios and the results show that the theoretical condition is satisfied.
Table 5.4: **Base Monte Carlo Option Prices**

This table presents prices for the lookback life annuity option with a time to maturity of three years. Prices for both men and women are expressed in percentages of the wealth at the issuance date of the option ($W_t$). We present option prices based on 90,000, 100,000 and 110,000 scenarios. Prices are significant up to less than 2 basispoints. The last row reports the average across all scenarios of the deflated cumulative stock returns.

<table>
<thead>
<tr>
<th></th>
<th>90,000</th>
<th>100,000</th>
<th>110,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>8.02</td>
<td>7.98</td>
<td>7.97</td>
</tr>
<tr>
<td>Women</td>
<td>9.22</td>
<td>9.17</td>
<td>9.17</td>
</tr>
<tr>
<td>$E[MR_{\pi,T}]$</td>
<td>0.997</td>
<td>0.999</td>
<td>0.995</td>
</tr>
</tbody>
</table>

### 5.4.2 Alternative Option Prices

We first determine the sensitivity of the option price with respect to the exercise price. An Asian option with the same time to maturity is priced. The Asian option holder has the right to convert his retirement wealth into a life annuity for the average interest rate during time to maturity.

Figure 5.4 shows that transforming the option into an Asian one, considerably lowers the price. For both men and women the price difference between the Asian and lookback option is approximately 6 percentage points. The Asian option has fewer scenarios with extremely high payoffs and many scenarios with a zero payoff. Only if the terminal rate $r_T$ is substantially lower than the average rate, high payoffs are realized.

As a second exercise price sensitivity test, we price an option that guarantees participants a certain conversion rate. Option prices are determined for guaranteed rates of 5% and 6%. Figure 5.4 again shows that the impact of changing the option version is large. Price differences between the guaranteed rate option and the base option range from 2.5 to 6 percentage points.

Beside the exercise price, we consider the sensitivity of the option price with respect to the time to maturity. The base option has a time to maturity of three years. We determine the effect of a change in time to maturity by pricing lookback life annuity option with times to maturity of 2, 4 and 5 years.

Figure 5.5 affirms the relatively large sensitivity with respect to the time to maturity documented in section 5.2. Decreasing the lookback period by one year leads to a price decrease of 2 percentage points, whereas an increase of one year causes a price increase of 2 percentage points. A lookback life annuity with a time to maturity of five years would cost a participant approximately 12% of the wealth at the option issuance date. Compared to the base option this implies a price increase of 4 percentage points.

Furthermore, we document the effect of a change in initial interest rate. The
Figure 5.4: Variation in Option Design

This figure displays the effect of a change in option design. Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women. Prices are provided for the base option (3 year lookback life annuity) and an Asian option and guaranteed rate option with equal characteristics. Conversion rates of 5% and 6% are guaranteed in the guaranteed rate option.
Figure 5.5: Variation in Time to Maturity
This figure displays the effect of a change in time to maturity. Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women. Prices are provided for the lookback life annuity option with times to maturity of respectively 2 year, 3 year (Base), 4 and 5 year.
5.4 Empirical Results

Figure 5.6: Variation in Initial Rate

This figure displays the effect of a change in initial conversion rate. Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women. Prices are provided for the base option (3 year lookback life annuity with initial rate of 4.42%) and for a lookback life annuity option with initial rates of 4%, 5% and 6% respectively.
base option has an initial rate equal to the last data point in our sample (4.42%). Figure 5.6 shows that the initial rate does not affect the option price. These findings support the proof in the appendix that the option payoff is not affected by changes in the minimum price.

As a final sensitivity test, we price the lookback life annuity option with different asset allocations. We restrict the investments during the lifetime of the option to one of the three available products: stock portfolio, two-year bond and ten-year bond.

A pure stock investment lowers the price compared to the base option by 1 percentage point. Restricting the investments to the two-year bond reduces the price by 4 percentage points. Since stocks have a higher risk premium than two-year bonds, the “stock-only” strategy is more expensive than the two-year bond strategy. However, when investments are restricted to the ten-year bond, we observe an increase in price. The price increase is a result of the risky character of the ten-year bond. Since we consider the ten-year bond rate as the conversion rate, restricting the investments to this category creates a multiplier effect. In relatively good scenarios this strategy leads to very high option payoffs via the difference in terminal and maximum interest rate and via high bond returns. In relatively bad scenarios this strategy leads to extremely poor payoffs vice versa. Nevertheless, we conclude that the asset mix drives the price of the option, but only to a minor extent.

5.5 Conclusion

Large numbers of uninformed plan participants face difficult decisions at retirement. Should they make an irreversible and risky investment in a life annuity or risk the possibility of outliving their money? If they buy a life annuity, when should they do so? These decisions affect their income for the rest of their lives. Especially the high value of timing (i.e. early retiring if the long-term interest rate is decreasing and postponing retirement in increasing interest rate states) the investment decision provokes regret aversion.

As a remedy for the regret aversion in that decision making process, we develop and price a lookback option on a life annuity. Participants buy this option in the last years before retirement. It provides them the right to lookback at retirement and buy a life annuity for the minimum forward price in the lookback period. First, we determine a closed-form lookback life annuity option price without the guarantee that the exact amount of retirement wealth is converted into a life annuity and under the assumption that the option holder invests risklessly during the time to maturity of the option. The closed-form option price is 7-9% of the plan participant’s wealth at the option issuance date. Then we price the option in an extended pension environment conditional on the guarantee that participants convert the exact amount of retirement wealth into a life annuity, with the possibility to invest in stocks and bonds during the time to maturity of the option and using a stochastic discount factor. The price for a defined contribution pension plan...
Figure 5.7: Variation in Asset allocation

This figure displays the effect of a change in asset allocation. Option prices are expressed in a percentage of wealth at the option issuance date, for both men and women. Prices are provided for the base option (3 year lookback life annuity with equal allocation in the three investment categories (stocks, two-year bond and 10 year bond)) and for a lookback life annuity option with investments restricted to one of the three categories.
plan participant to insure against regret aversion at retirement, whilst retaining investment freedom, is 8-9% of the wealth three years before retirement. Hence, closed-form and simulated prices are comparable.

In the closed-form setting we document that interest rate volatility and time to maturity are important option price drivers. In the simulation setting, we determine the option Greeks by pricing alternative options. Another important driver of the option price is the lookback feature (i.e. the exercise price). If participants are offered the right to buy the life annuity for the average forward price (Asian option) or for a predetermined price, the option price decreases substantially. We find that the option price is neither sensitive to the initial interest rate, nor to the asset allocation during the time to maturity.

Future research could extend our analysis in multiple ways. Participants could be offered real instead of nominal pension income. Furthermore, mortality risk and life annuity risk premia could be added in the option pricing environment.

5.A Appendix

This appendix proves that the hedging strategy as specified in section 5.2 is a replicating and self-financing strategy for a lookback option on a discount bond. To formally prove that the hedging portfolio is indeed replicating and self-financing, we check the following two conditions (see Baxter and Rennie, 1996):

**Condition 5.A.1.**

If \((\phi^T_t, \phi^s_t)\) is a portfolio and \((P(r,t,T), P(r,t,s))\) are the time \(t\) prices of bonds maturing at \(T\) and \(s\) respectively, then \((\phi^T_t, \phi^s_t)\) is **self-financing** if and only if \(dH = \phi^T_t dP(r,t,T) + \phi^s_t dP(r,t,s)\).

**Condition 5.A.2.**

A **replicating strategy** for \(U_T\) is a self-financing portfolio \((\phi^T_t, \phi^s_t)\) such that

\[
H(r,t,Q) = \phi^T_t P(r,t,T) + \phi^s_t P(r,t,s) = U_T.
\]

It is straightforward to check that the payoff of the hedging portfolio replicates the option payoff \(U_T\). When the maturity date of the option is approaching, equations (5.5) to (5.10) show that the portfolio has limit value

\[
\lim_{t \to T} H(r,t,Q) = P(r,T,s) - Q \left( f_T, t_a, T, s \right),
\]

which is identical to the option payoff as specified in equation (5.2).

To prove that the hedging portfolio is self-financing, we have to prove that

\[
dH = \phi^T_t dP(r,t,T) + \phi^s_t dP(r,t,s).
\]

An application of Itô’s lemma to expression (5.5) shows that changes in hedging portfolio value can be characterized as

\[
dH = H_r dr + H_Q dQ - H_t dt + \frac{1}{2} H_{rr} (dr)^2.
\]
Since the hedging portfolio is a function of the time to maturity \( T - t \), the dependence of the hedging portfolio on \( t \) is negative, as indicated by the minus in equation (5.35).

We start with proving that \( H_Q dQ = 0 \). To prove this we can restrict ourselves to the cases where the forward bond price \( F^B(r, t, T) = \frac{P(r, t, s)}{P(r, t, T)} \) attains a minimum

\[
F^B(r, t, T) \rightarrow Q(\bar{f}_t, t, s).
\]

If the forward bond price is not at a minimum \( dQ = 0 \) and hence \( H_Q dQ = 0 \). We therefore have to prove that \( H_Q = 0 \) if the forward bond price attains a minimum. The methodology to prove this is introduced by Goldman, Sosin, and Gatto (1979). We prove that the moments of \( Q(\bar{f}_T, t, T, s) \), conditional on \( P(r, t, s) \), \( t \) and \( Q(\bar{f}_t, t, s) \) are independent of \( Q(\bar{f}_t, t, s) \) for any \( t < T \). Since the hedging portfolio payoff and therefore also its value is a function of \( Q(\bar{f}_T, T, s) \), it suffices to show that the distribution of \( Q(\bar{f}_T, T, s) \) is unaffected by changes in \( Q(\bar{f}_t, t, s) \).

To see the dependence of the option price on \( Q(\bar{f}_T, T, s) \) we note that

\[
C^B_L(P(r, t, s), Q(\bar{f}_t, t, s), T - t, s) = e^{-rt} E[P(r, T, s) - Q(\bar{f}_T, T, s)]. \tag{5.36}
\]

Let

\[
\psi = \begin{cases} 
\frac{Q(\bar{f}_t, t, s)}{F^B(r, t, T, s)} & \text{if } Q(\bar{f}_t, t, s) < Q_T(\bar{f}_T, t, T, s) \\
\frac{Q_T(\bar{f}_T, t, T, s)}{F^B(r, t, T, s)} & \text{otherwise,}
\end{cases} \tag{5.37}
\]

and let \( Q_T(\bar{f}_T, t, T, s) \) and \( f_{t,T} \) respectively be the minimum bond price and maximum forward rate in the period \([t, T] \). By construction \( Q(\bar{f}_T, T, s) = F^B(r, t, T, s)\psi \).

The \( n^{th} \) moment of \( Q(\bar{f}_T, T, s) \) is defined as

\[
G(n) = \int_{-\infty}^{\infty} [Q(\bar{f}_T, T, s)]^n d\Phi_Q [Q(\bar{f}_T, T, s)], \tag{5.38}
\]

with \( \Phi_Q \) the conditional CDF of \( Q(\bar{f}_T, T, s) \) given \( F^B(r, t, T, s) \). Substitution transforms the \( n^{th} \) raw moment into

\[
G(n) = \int_0^1 [F^B(r, t, T, s)\psi]^n d\Phi_Q[\psi]. \tag{5.39}
\]

Decomposing \( \psi \) as in the definition and noting that we condition on the information set at time \( t \) leads to

\[
G(n) = \left[ Q(f_t, t, s)^n \int_{Q(\bar{f}_t, t, s)}^{1} d\Phi_Z(Z) \right] + \left[ Q(f_t, t, s)^n \int_0^{Q(\bar{f}_t, t, s)} Z^n d\Phi_Z(Z) \right],
\]

125

Thesis_Rik V1.pdf
with \( \Phi_Z \) the distribution function of \( Z \). \( Z \) is like a return and \( F^B(r, t, T, s) \) follows a random walk (see Jamshidian, 1989). Therefore the derivative of the second part towards \( Q(\tilde{f}_t, t, s) \) is zero. Differentiation of the first part with respect to \( Q(\tilde{f}_t, t, s) \) yields

\[
\frac{\partial G(n)}{\partial Q(\tilde{f}_t, t, s)} = nQ(\tilde{f}_t, t, s)^{n-1} \int_{\frac{Q(\tilde{f}_t, t, s)}{\Phi_{\tilde{f}_t}(r, t, s)}}^1 d\Phi_Z(Z). \tag{5.41}
\]

Since we only consider the case where \( F^B(r, t, T, s) \to Q(\tilde{f}_t, t, s) \) and we assumed an Ornstein-Uhlenbeck process for the underlying interest rate, the probability measure at the point \( \frac{Q(\tilde{f}_t, t, s)}{\Phi_{\tilde{f}_t}(r, t, s)} \) is zero. Hence, the \( n \)-th moment of \( Q(f_T, T, s) \) is independent of \( Q(\tilde{f}_t, t, s) \) for any \( n \) and therefore \( H_Q = 0 \) if the forward bond price attains a minimum.

After proving that \( H_QdQ = 0 \) and substituting equation (5.1) for \( dr \), the change in hedging portfolio value is characterized as

\[
dH = H_r \sigma_r \, dW + \left[ H_r \alpha (\gamma - r) + \frac{1}{2} H_{rr} \sigma_r^2 - H_t \right] \, dt. \tag{5.42}
\]

Differentiating the hedging portfolio value twice towards \( r \), and towards \( t \) leads to:

\[
H_r = -\phi_t^1 \frac{1 - e^{-\alpha(s-t)}}{\alpha} P(r, t, s) - \phi_t^T \frac{1 - e^{-\alpha(T-t)}}{\alpha} P(r, t, T), \tag{5.43}
\]

\[
H_{rr} = \phi_t^1 \left( 1 - e^{-\alpha(s-t)} \right)^2 \frac{1 - e^{-\alpha(T-t)}}{\alpha^2} P(r, t, s) + \phi_t^T \left( 1 - e^{-\alpha(T-t)} \right)^2 \frac{1}{\alpha^2} P(r, t, T)
- 2n(h)P(r, t, s) \frac{1}{\sigma_p \alpha^2} \left[ e^{-2\alpha(T-t)} \left( 1 - 2e^{-\alpha(s-T)} + e^{-2\alpha(s-T)} \right) \right], \tag{5.44}
\]

\[
H_t = -\phi_t^1 P(r, t, s) \left( e^{-\alpha(s-t)} \gamma + (1 - e^{-\alpha(s-t)}) \gamma - \sigma_r^2 \frac{1 - e^{-\alpha(s-t)}}{2\alpha^2} \right)
- \phi_t^T P(r, t, T) \left( e^{-\alpha(T-t)} \gamma + (1 - e^{-\alpha(T-t)}) \gamma - \sigma_r^2 \frac{1 - e^{-\alpha(T-t)}}{2\alpha^2} \right)
- n(h)P(r, t, s) \frac{\sigma_r^2}{\sigma_p \alpha^2} e^{-2\alpha(T-t)} \left[ 1 - 2e^{-\alpha(s-T)} + e^{-2\alpha(s-T)} \right]. \tag{5.45}
\]
Substitution of the expressions derived for $H_t$, $H_r$, and $H_{rr}$ gives

\[
dH = \left[ \phi_t^{(T)} P(r, t, T) + \phi_t^{(s)} P(r, t, s) \right] rdt \tag{5.46}
\]

\[
- \left[ \phi_t^{(T)} P(r, t, T) \frac{1 - e^{-\alpha(T-t)}}{\alpha} + \phi_t^{(s)} P(r, t, s) \frac{1 - e^{-\alpha(s-t)}}{\alpha} \right] \sigma_r dW. \tag{5.47}
\]

\[
= \phi_t^{(T)} dP(r, t, T) + \phi_t^{(s)} dP(r, t, s) \tag{5.48}
\]

Hence the hedging portfolio $H_t$ is self-financing and consequently a replicating portfolio.
Chapter 6

Conclusion

This thesis addresses financial risk management, first from a global, macroeconomic perspective and thereafter narrowing it down per chapter to the individual pension plan participant. As a main overall conclusion, we can state that it is crucial for risk management purposes to accurately quantify investment risks. Better risk measurement enhances risk management decisions. We suggest several different approaches to improve risk measurement, ranging from more complete data sets to improved model specifications and estimation techniques. All of our methods contribute to the risk management of institutional investors and private individuals. Since the diversity of the topics mentioned in this thesis is too broad to draw more general conclusions, we proceed on a per chapter basis.

In chapter 2 we show that expectations about the Atlantic trade and speculation in insurance companies may have been important factors driving share prices in the turbulent year 1720. More specifically, long-term prospects for Atlantic commerce, and financial innovation in the organizational form of insurance companies may have induced a bubble in international asset prices. Our cross-sectional analysis shows that shares of companies trading with the Americas or involved in the insurance industry had price run-ups exceeding other industries by large amounts. We also show that Atlantic trading firms experienced a permanent price increase, while firms in other industries had mixed results. Our cross-sectional evidence suggests that the “bubble” may have been based upon some fundamental common factor that justified a value increase. While investor irrationality may have carried prices to many multiples of their post-crash value, the bubble speculation may have anticipated some long-term permanent effect.

In chapter 3 we improve both the specification and estimation of firm-specific, time-varying equity betas. We combine parametric and non-parametric approaches for modeling changes in betas. The precision of firm-level beta estimates is increased by setting up a Bayesian panel data model that exploits information contained in the cross-section of stocks and imposes a common structure on parameters while still allowing for cross-sectional heterogeneity. Fundamental and realized betas exhibit different time series dynamics and cross-sectional characteristics, and
therefore a combination of these specifications captures more aspects than either of the separate specifications. The optimal mixture of the two beta components varies across firms and over time. We further demonstrate that our panel data approach yields more precise estimates of firm-level betas than the traditional approach of estimating betas by running a time series regression for every firm. Moreover, we document strong cross-sectional variation in betas of firms that are grouped together in traditional asset pricing portfolios. Consequently, aggregating stocks into portfolios conceals important information contained in individual stock betas and reduces the cross-sectional variation in betas. We also demonstrate that the mixed betas generated by our panel data model lead to a sharp increase in the pricing ability of the conditional CAPM. Furthermore, we support the finding of Ang, Liu, and Schwarz (2008) that the use of individual stocks as tests assets instead of portfolios leads to more efficient estimates in cross-sectional tests of asset pricing models. We extend their work by showing that a better specification and more precise estimation of stock-specific betas increases the explanatory power of the CAPM. Additionally, we use our beta estimates for forecasting the covariance matrix out-of-sample in a minimum variance optimization. The out-of-sample standard deviation of portfolios constructed using our betas is lower than that of competing methods.

In chapter 4, we document the domestic equity performance of US pension funds. We employ a data set that encompasses a large part of the US pension industry, multiple plan types, various mandate types and contains vast information on returns, costs and benchmarks. We provide insights in the average net risk-adjusted fund return, average cost levels and benchmarks used to evaluate asset managers. We also test for possible biases that result from the voluntary reporting of pension funds to our data provider (CEM) and find none. We find that pension funds' net returns are very close to standard benchmarks, cost levels are substantially lower than standard mutual fund fees and are strongly diminishing in fund size. Additionally, we give an overview of the benchmarks used by pension funds to evaluate their asset managers and show that pension funds use many different benchmarks both across mandates and over time. We argue that pension funds have bargaining power with external parties, scale advantages in internally managed portfolios and monitoring ability, which does not lead to higher returns, but reduces the amount and likelihood of underperformance.

In chapter 5 we develop and price a lookback life annuity option, which protects defined contribution plan participants against annuity risk at retirement and provides protection against regret in the retirement timing decision. It grants participants the right to look three years back at retirement and buy a life annuity for the minimum forward price in the lookback period. We determine a closed-form lookback life annuity option price without the guarantee that the exact amount of retirement wealth is converted into a life annuity and under the assumption that the option holder invests risklessly during the time to maturity of the option. Thereafter, we price the option in an extended pension environment conditional on the guarantee that participants convert the exact amount of retirement wealth into a life annuity, with the possibility to invest in stocks and bonds during the
time to maturity of the option and using a stochastic discount factor. The closed-
form option price is 7-9% of the plan participant’s wealth at the option issuance
date. The price for a defined contribution plan participant to insure against regret
aversion at retirement, whilst retaining investment freedom, is 8-9% of the wealth
three years before retirement. Hence, differences between closed-form and simu-
lated prices are negligible. In the closed-form setting we document that interest
rate volatility and time to maturity are important option price drivers. In the
simulation setting, we determine the option Greeks by pricing alternative options.
Another important driver of the option price is the lookback feature (i.e. the ex-
ercise price). If participants are offered the right to buy the life annuity for the
average forward price (Asian option) or for a predetermined price, the option price
decreases substantially. We finally find that the option price is neither sensitive
to the initial interest rate, nor to the asset allocation during the time to maturity
of the option.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


138

Mercer, 2006, Global Investment Management Fee Survey.


Milevsky, M., and V. Young, 2006, Annuitization and asset allocation, .


BIBLIOGRAPHY


Van Rijn, G., 1899, *De Actiehandel in 1720 te Rotterdam en de Maatschappij van Assurantie, Discontering en Beleening dezer Stad* (P. M. Bazendijk: Rotterdam).


Het meten, controleren en beheersen van financieel risico is van cruciaal belang voor het functioneren van onze maatschappij. Vaak wordt gedacht dat financieel risico vooral belangrijk is voor grote ondernemingen en wordt er bij financiering vooral gedacht aan de effectenbeurs. Echter, financiering en financieel risico bepalen in grote mate ons welvaartsniveau en verklaren in veel gevallen waarom bepaalde beslissingen (soms achter de schermen) genomen worden. Zoals Robert Shiller al aangeeft in zijn eerste college aan nieuwe Yale studenten, leer je de wereld begrijpen aan de hand van financiering. Hij zegt dat iemand misschien wel zal denken: "Ik ben een dichter, wat heb ik met financiering te maken?". Maar voordat je het weet zit je aan tafel bij je uitgever en wordt er over financiën gesproken.

Dit proefschrift draagt bij aan het meten en beheersen van financieel risico. We dragen op verschillende manieren bij aan dit onderwerp. We starten op een globaal niveau door internationale zeepbellen in aandelenprijzen te bestuderen en vernauwen ons gezichtsveld telkens verder.

In hoofdstuk 1, leiden we het onderwerp in en benadrukken we het belang van risicobeheersing voor zowel institutionele beleggers als individuen. We zetten ook uiteen waarom het van cruciaal belang is dat we financieel risico nauwkeurig kunnen meten.

In hoofdstuk 2 bekijken we welke factoren ertoe bijgedragen hebben dat veel investeerders in 1720 besloten aandelen te kopen en vervolgens veel geld verloren. Vaak worden deze beslissingen toegeschreven aan latente, publieke hebzucht of wereldwijd irrationeel of ongefundeerd enthousiasme. Wij proberen nieuwe verklaringen voor de zeepbellen te vinden door de verschillen in grootte tussen de bellen van diverse aandelen te bestuderen.

Zo laten we zien dat aandelen van ondernemingen die handelden met het Westen (Royal African Company, WIC en South Sea Company) veel sterker stegen dan aandelen van ondernemingen die geen handel dreven met het nieuw-ontdekte gebied. Wat zou de onderliggende oorzaak kunnen zijn van het vertrouwen in de ondernemingen die met Amerika handelden? Wij vermoeden dat de afnemende Spaanse dominantie in het Westen een belangrijke rol gespeeld heeft. Spanje had het recht om in het Westen slaven te verhandelen (Asiento) verloren en had in de laatste twee oorlogen (Spaanse Successieoorlog en de Oorlog van de Quadruple Alliantie) veel gebied aan Engeland (Groot-Britannië) verloren. Ook Frankrijk
verloor veel gebied aan Groot-Brittanië. Wellicht dat de recente overwinningen tot optimisme onder de Britse belegger geleid hebben.

Ook tonen we aan dat aandelen van nieuwe verzekerringsmaatschappijen veel sterker in prijs stegen dan de overige aandelen. Rond het jaar 1720, vond er een belangrijke financiële innovatie plaats en waren verzekerringsmaatschappijen voor het eerst in staat om kapitaal te verzamelen door de uitgifte van aandelen. De aandeelhouder kon zo slechts zeer beperkt afhankelijk worden van de financiële situatie van de onderneming. Dit was een enorme vooruitgang ten opzichte van de grote afhankelijkheid die investeerders voordien hadden. Verzekerringen werden voor de innovatie namelijk gefinancierd door kapitaalkrachtige handelaren die grote sommen geld verenigden. De kapitaalkrachtigen waren zo met grote sommen geld afhankelijk van de onderneming. Veel beleggers zagen de toegevoegde waarde van de nieuwe ondernemingsvorm in en hoopten mee te profiteren van de toekomstige winsten door aandelen aan te schaffen.

In hoofdstuk 3 stellen we een nieuwe manier voor om het risico van individuele aandelen te kwantificeren. We brengen een verbetering aan in de modelspecificatie die gebruikt worden om risico te kwantificeren, en we passen nieuwe, efficiëntere schattingsmethoden voor deze modellen voor. De verbeteringen leiden tot een grotere verklaringskracht in cross-sectionele testen van aandelenprijzen, ten opzichte van concurrerende methoden.

Een van de meest gebruikte methoden om het risico van een individueel aandeel weer te geven, is de CAPM-beta. Hoewel verschillende studies aangetoond hebben dat deze beta tijdsvariërend is, bestaat er grote oneenigheid over de manier waarop deze tijdsvariatie gemodelleerd dient te worden. Aan de ene kant zijn er studies die de tijdsvariatie modelleren door beta afhankelijk te maken van economische factoren. Aan de andere kant, is er een andere groep wetenschappers die kiest voor een data-gedreven beta, bijvoorbeeld door beta’s op korte intervallen te schatten of beta autoregressief te maken. Wij combineren beide methoden en laten het gewicht dat aan elk van de methoden gegeven wordt per aandeel en door de tijd variëren.

Daarnaast verbeteren we schattingsmethoden door een hiërarchisch Bayesi- aans panel model toe te passen. Door een gelijke verdeling te veronderstellen voor fonds-specifieke parameters, is het model flexibel, maar zijn de schattingen toch erg nauwkeurig. Het model heeft de flexibiliteit om veel gewicht aan fonds-specifieke schattingen te geven wanneer deze accuraat zijn en minder wanneer ze minder nauwkeurig zijn. Ook maken we gebruik van hoog-frequentie data in de schatting van de data-gedreven beta, waarbij we een gewichtenfunctie specificeren die meer gewicht geeft aan meer recente observaties. Deze verbeteringen leiden tot een grote toename in de nauwkeurigheid van de beta schattingen, gemeten in betrouwbaarheidsintervallen.

In een toepassing laten we zien dat onze verbeterde beta schattingen ook leiden tot een grotere verklarende kracht in cross-sectionele aandelenprijzingstochten. We laten zien dat de risicopremie dichter bij het datagemiddelde ligt, dat de $R^2$ van ons model aanzienlijk hoger is dan bij alternatieve modellen en dat ons model een significant positieve lading op de fonds-specifieke beta’s heeft. Een andere
Nederlandse Samenvatting

toepassing laat zien dat er aanzienlijke verschillen in beta’s bestaan binnen traditionele portefeuilles die gebruikt worden voor aandelenprijzingsstoetsen. Dit is in strijd met de populaire aanname dat beta’s binnen een dergelijke portefeuille gelijk zijn. Tot slot tonen we aan dat onze nauwkeurigere beta schattingen voordeel opleveren bij de constructie van portefeuilles met minimum variantie. In een out-of-sample analyse verslaan we met behulp van onze beta’s concurrerende modellen (bijvoorbeeld tijdreeks-geschatte beta’s) in termen van standaarddeviatie.

In hoofdstuk 4 zoomen we verder in door enkel naar de binnenlandse aandelenportefeuilles van Amerikaanse pensioenfondsen te kijken. We meten de gemiddelde rendementen na risicocorrectie. Ook meten we de gemiddelde kosten niveaus en schaalvoordelen in kosten van grote pensioenfondsen. Verder geven we aan welke indices vaak door pensioenfondsen gebruikt worden als referentiepunt voor beoordeling van de aandelenrendementen (referentie-indices).

Door een gebrek aan data over pensioenfondsen, bestaat er geen consensus over de prestaties van hun aandelenportefeuilles. Veel gebruikte databases bevatten slechts een bepaald type fonds, data op portefeuilleniveau en niet op fondsniveau. Bovendien zijn de data onderhevig aan biases omdat pensioenfondsen nu eenmaal geen verplichting hebben om hun resultaten periodiek te rapporteren. Gezien de grote diversiteit tussen de verschillende databases, is het niet vreemd dat verschillende studies tot uiteenlopende bevindingen komen. Verder is data over de kosten van pensioenfondsen of referentie-indices vrijwel niet-bestaand.

Wij geven inzicht in de prestaties van de binnenlandse aandelenportefeuilles van Amerikaanse pensioenfondsen, waarbij we gebruik maken van een veelomvattende database. Onze data lijdt niet aan eerder genoemde biases, en bevat uitgebreide informatie over rendementen, kosten en referentie-indices. Ook bevatten onze database vele soorten pensioenfondsen en hebben wij data op fondsniveau. We laten zien dat Amerikaanse pensioenfondsen, na risicocorrectie, rendementen hebben die heel dicht bij de referentie-indices liggen. Verder tonen we aan dat de kosten niveaus van pensioenfondsen een stuk lager zijn dan van beleggingsfondsen en dat er grote schaalvoordelen in kosten zijn. Daarnaast geven we inzichten in de indices die pensioenfondsen als referentie gebruiken (referentie-indices). We laten zien dat er een breed scala aan indices gebruikt wordt voor referentie-doeleinden en dat de indices per fonds en over de tijd veranderen. We tonen aan dat pensioenfondsen beter dan individuele beleggers of beleggingsfondsen in staat zijn kosten te reduceren. Verder beweren we dat pensioenfondsen door hun grootte beter in staat zijn om hun portefeuilles te bewaken, een grote onderhandelingskracht hebben met externe partijen en interne schaalvoordelen hebben. Dit leidt niet tot hogere rendementen, maar stelt hen wel in staat om de kans op en grootte van slechte rendementen te beperken.

In hoofdstuk 5 verkleinen we onze focus nog verder en bestuderen we de risico’s van pensioeneelenners in een beschikbaar premiesysteem. Vlak voor de pensioenedatum dient een deelnemer zijn pensioen om te zetten in een lijfrente. Hierbij wordt hij blootgesteld aan veel renterisico en loopt hij een groot risico om spijt te krijgen van het moment waarop hij zijn vermogen geconverteerd heeft. Een half jaar eerder of later met pensioen gaan kan namelijk vrij eenvoudig resulteren in

143
inkomensverschillen van 10%. Wij ontwikkelen en waarderen een optie die hem beschermt tegen beide risico’s.

Wij ontwikkelen en prijzen een lookback optie op een lijfrente. Het lookback-karakter van de optie biedt de koper bescherming tegen mogelijke spijt van het kiezen van de pensioneringsdatum. Verder beschermt de optie de deelnemer ook tegen het renteisico dat hij loopt bij de conversie van zijn opgebouwde kapitaal in een lijfrente. We leiden eerst een closed-form optieprijs af, onder restrictieve aannames met betrekking tot de beleggingsmogelijkheden gedurende de looptijd van de optie. Daarna laten we enkele restrictieve aannames varen en waarderen de optie met behulp van simulatietechnieken.

We tonen aan dat de prijs van de optie ongeveer 8% van het opgebouwde kapitaal bedraagt. Ook laten we zien dat de hoogte van de prijs vooral bepaald wordt door het lookback-karakter van de optie, i.e. de prijs van bescherming tegen spijt is relatief hoog. Wanneer we het lookback-kenmerk verwijderen, wordt de optie aanmerkelijk goedkoper. Ook toont de optieprijs gevoeligheid met betrekking tot de volatiliteit van het onderliggende renteproces en de looptijd van de optie. Echter, de beleggingsmix en de rentestand bij aankoop van de optie spelen geen belangrijke rol in de waardering van de optie.

In hoofdstuk 6 geven we een overzicht van de belangrijkste conclusies van dit proefschrift.
Rik Frehen was born on March 18, 1982 in Heerlen, The Netherlands. He attended Eijkhagen College in Landgraaf between 1994 and 2000. Subsequently, he studied Econometrics at Maastricht University. In September 2004 he obtained his master’s degree.

After graduation, Rik taught business economics at Eijkhagen College in Landgraaf. In July 2005, he joined the Finance Department of Maastricht University as a Ph.D. candidate. Rik conducted part of his research while attending Yale University as a post-doctoral fellow. The results of his research are presented in this thesis.


In October 2009, Rik joined the Finance Department of Tilburg University as an assistant professor.