Essays on Risk Assessment and Risk Management

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Management

submitted by

Thomas Rolf Althaus
from
Walkringen (Bern)

Approved on the application of

Prof. Paul Söderlind, PhD

and

Dr. Ralf Seiz

Dissertation no. 4323
ZSUZ Druckerei, Zürich 2015

The University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, October 22, 2014

The President:

Prof. Dr. Thomas Bieger

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Paul Söderlind and co-supervisor Dr. Ralf Seiz who have supported me with invaluable advice when writing this thesis. They have allowed me to explore the topics I was most interested in and guided me on how to convert my ideas into this thesis for which I am deeply grateful.

A very special thank-you is dedicated to Olive who has accompanied me on this journey and has always supported my endeavours with loving care, especially during the tougher times. I am also very grateful to my parents, Kathrin and Rolf, who have provided me with a multitude of opportunities in life and in particular enabled my studies at the university. Moreover, I would also like to thank my sister Martina and my brother Alexander for providing me with plenty of carefree hours spent together, even during otherwise stressful episodes. Also, I would like to thank all the close friends who have accompanied and supported me for the past years and look forward to the time we will spend together in the future.

Zürich, November 2014 Thomas Althaus

Contents

List of Tables

List of Figures

Executive Summary

This thesis consists of three singled-authored essays which examine topics in risk assessment and risk management in three distinct fields. The first essay, presented in chapter 1, introduces a new set of risk premia based factors which are used to explain the performance of hedge funds. It is shown that these factors are superior at explaining the average variation as well as the level of hedge fund returns in different hedge fund categories when compared to a benchmark factor model, using the option based straddle factors introduced in Fung and Hsieh (2001). Further, the new risk factors are applied in a multiple structural change regression approach. The results find evidence that single hedge funds tend to switch their exposure to the risk factors at points in time, when large shifts in the financial markets take place and that the percentage of hedge funds generating significant positive alpha has decreased over. The second essay, presented in chapter 2, compares the performance of the different risk models at predicting the Value-at-Risk, the Expected-Shortfall as well as predicting the entire left-tail of the return distribution for different static and dynamic indices. The results indicate that methods based on a fully parametric approach using a skewed-t distribution respectively a generalized asymmetric student-t distribution to model the innovations perform the best. The third essay, presented in chapter 3, concludes by examining the question whether common portfolio insurance strategies such as constant proportion portfolio insurance strategies, stop-loss strategies or synthetic put replication strategies are beneficial in the context of a pension fund framework. Thereby, a proxy of the Swiss pension system is modeled in a simulation framework, where the employee's as well as the employer's preferences are tracked. The results indicate that dynamic portfolio insurance strategies may be preferred over passive strategies depending on the target return of the pension fund, the long term interest rate assumptions as well as the approach taken to model the employee's preferences as well as the employer's preferences.

Zusammenfassung

Diese Dissertation besteht aus drei Aufsätzen, welche Themen aus dem Bereich Risiko-Beurteilung und Risiko-Management behandeln. Der erste Aufsatz in Kapitel 1 führt neue risikoprämienbasierte Faktoren ein, welche benutzt werden um die Performance von Hedge Funds zu erklären. Es wird gezeigt, dass diese Faktoren die durchschnittliche Variation sowie die Höhe von Hedge Fund Renditen besser erklären können als ein Benchmark-Modell, welches die optionsbasierten Straddle Faktoren von Fung and Hsieh (2001) benutzt. Des Weiteren werden die neuen Risikofaktoren in einem multiple structural change Regressionsmodell angewandt. Die Resultate dazu zeigen, dass einzelne Hedge Funds das Exposure zu den Risikofaktoren häufig dann wechseln, wenn grosse Verschiebungen an den Finanzmärkten stattfinden. Zudem weisen die Ergebnisse darauf hin, dass der Prozentsatz von Hedge Funds, welche ein signifikant positives Alpha generieren über die Zeit abgenommen hat. Der zweite Aufsatz in Kapitel 2 vergleicht die Leistungsfähigkeit von verschiedenen Risikomodellen in Bezug auf die Schätzung von Value-at-Risk und Expected-Shortfall Risikomassen für statische sowie dynamische Indizes. Die Ergebnisse zeigen, dass Risikomodelle, welche auf einem parametrischen Ansatz beruhen, der die schiefe studentsche t-Verteilung oder die generalisierte asymmetrische studentsche t-Verteilung verwenden, die besten Resultate erzielen. Der dritte Aufsatz in Kapitel 3 untersucht abschliessend die Frage, ob Portfolio-Absicherungsstrategien im Kontext einer Pensionskasse bessere Nutzen-Risiko Resultate erzielen können als passive Strategien. Dazu wird eine Simulationsanalyse durchgeführt, welche auf dem Schweizer Pensionskassensystem basiert und es werden sowohl die Präferenzen der Arbeitnehmer sowie des Arbeitgebers berücksichtigt. Die Ergebnisse zeigen, dass Portfolio-Absicherungsstrategien passive Strategien unter gewissen Umständen dominieren. Dabei spielen die Zielrenditen der Pensionskasse, die Annahmen über das Zinsniveau, sowie die Definition der Arbeitnehmer- und Arbeitgeber-Präferenzen eine entscheidende Rolle.

Chapter 1

Hedge Fund Risk Dynamics: Introducing New Risk Factors

1.1 Introduction

There exists a broad range of literature which is focused on explaining the return characteristics and performance of hedge funds by using linear multi-factor models[1]. In this context Fung and Hsieh (2001) introduce risk factors based on option portfolios (lookback straddle portfolios) in five asset classes, which are used to explain the returns of trend following funds. Further, in Fung and Hsieh (2004) they propose a seven factor model which includes three of the option portfolio based risk factors to explain the returns of hedge fund indices[2]. This seven factor model is for example able to explain between 55% and 80% of the monthly return variation of the HFR Fund of Funds Index for the period of January 1994 to December 2002. Many subsequent studies focused on hedge fund performance measurement use the factors proposed by Fung and Hsieh (2004) for their analysis, whereby the

[1]The first study investigating the performance of mutual funds goes back to Jensen (1968).

[2]The seven factors are the S&P 500 index, a small cap factor (Wilshire Small Cap 1750 index - Wilshire Large Cap 750 index), a bond factor (month-end to month-end change in the U.S. Federal Reserve 10-year constant-maturity yield), a credit factor (month-end to month-end change in the difference between Moody's Baa yield and the Federal Reserve's 10-year constant-maturity yield, and three option factors (returns of a portfolio of lookback straddles on bond, currency and commodity futures respectively). A list of the factors which are used in this study as the benchmark model is found in table 1.2.

set of factors is sometimes extended by additional new factors or the factors introduced by Agarwal and Naik (2004)[3]. Therefore, the factor model of Fung and Hsieh (2004) is often regarded as a benchmark model. Examples of recent research relying on the Fung and Hsieh (2001) option factors are Bollen and Whaley (2009), Bali, Brown and Caglayan (2011), Huber (2011), Titman and Tiu (2011), Buraschi, Kosowski and Trojani (2011) and Bollen (2011). Despite the widespread use of the benchmark model with its option based factors, this approach has shortcomings which have been addressed in the existing literature.

A first criticism which has been raised, is related to the fact that the option based straddle factors are not able to properly explain the level of hedge fund returns. In this context, Gisclon (2011) shows that the Fung and Hsieh (2001) straddle based risk factors are able to explain the time-variation of CTA funds well, while overestimating the alpha. He shows that dynamic benchmark factors, which correspond more closely to the trading practices of CTA funds, are superior at explaining the variation as well as the level of the excess returns of CTA funds compared to the Fung and Hsieh straddle factors. The reason for the poor performance of the straddle factors at explaining the alpha of CTA funds lies in the risk premia for these factors which are either negative or close to zero. He concludes, that CTA funds should be modeled in a linear factor model framework, using dynamic benchmark factors. A second criticism of the benchmark model, which has been raised for example in Bollen and Whaley (2009), is the notion that the factors proposed in the model are not easily tradeable and therefore do not represent returns which could be achieved in a practical real world setting.

[3]Agarwal and Naik (2004) introduce a similar risk model where they use 12 "buy-and-hold risk factors" and 4 equity option-based risk factors. The buy and hold factors are: Russell 3000 index, MSCI world ex US index, MSCI emerging markets index, Fama and French's (1993) SMB factor, Fama and French's (1993) HML factor, Carhart's (1997) momentum factor, Salomon Brothers government and corporate bond index, Salomon Brothers world government bond index, Lehman high yield index, Federal Reserve Bank competitiveness-weighed dollar index, Goldman Sachs commodity index, and the change in the default spread in basis points between corporate and government bonds. The option based factors are portfolios of options on the S&P 500 index which are rolled over every month.

Following this critique, this study will introduce new risk factors which are based on static and dynamic strategies for which there is empirical evidence that they carry a positive risk premium over long term horizons and which are based on liquid tradeable instruments. The approach to defining risk factors in this study is inspired by the research of Asness, Moskowitz and Pedersen (2013) who analyze the returns of momentum and value premia in different asset classes and Bender et al. (2010) who introduce the idea of building portfolios not by choosing asset classes, but by choosing risk premia within different asset classes[4]. Since the used strategies, which can be interpreted as risk premium based factors, show favorable diversification properties as shown in Asness, Moskowitz and Pedersen (2013) and Bender et al. (2010), the idea of this study will be to analyze how well such factors perform in a multi-factor regression framework versus the benchmark multi-factor model at explaining the performance of hedge funds. If these risk premium based factors are superior at describing the performance of hedge funds, they pose a viable alternative to the benchmark factors, particularly since they are easily tradeable[5]. Thereby, the multifactor model is seen as a performance attribution model and not as an asset pricing model[6].

---

[4] Asness, Moskowitz and Pedersen (2013) demonstrate that value and momentum effects exist not only in equities but also in bonds, commodities and currencies. They show that value and momentum strategies are negatively related within and across asset classes, which implies that value and momentum strategies have a positive effect on diversification when combined in a portfolio. Further, they find some evidence that momentum and value effects are related to liquidity risk. Bender et al. (2010) go one step further, by building a portfolio of value and momentum strategies together with other strategies for which they expect to capture a premium. They show that such a portfolio has superior risk return characteristics compared to a common portfolio existing only of bonds and equities.

[5] Institutional investors are able to easily gain exposure to strategies mimicking the new risk factors described in this study. An overview of strategies which are available can be found for example on the following homepage of Barclays Capital: https://ecommerce.barcap.com/indices/index.dxml

[6] Bollen and Whaley (2009) refer to Carhart (1997) who distinguishes interpreting a multi-factor model either as a model of market equilibrium or a performance attribution model.

Considered New Risk Factors

The risk factors considered in this study are motivated by academic research which finds evidence that there is a risk premium related to a specific factor or that the strategy defining the factor has beneficial diversifying properties in a portfolio context. Also, the type of strategy should be available in practice, whereby I refer to the offering of Barclays Capital as a proxy[7]. However, the factors do not have to be motivated by an asset pricing model and no claim is made that the following choice of strategies is complete.

Likely the most prominent research piece on risk factors is Fama and French (1993), which suggest that there are three common stock market factors (market, size, value) and two bond-market factors (term risk, credit risk). A factor closely related to the equity factors introduced by Fama and French (1993) is an equity momentum factor as defined by Jegadeesh and Titman (1993). These factors have at least partially been included in the hedge fund literature. However, there is strong evidence that there exist other common risk factors which have not gained much attention in the context of the hedge fund literature. One example is a risk factor related to the forward premium of short term interest rates. While there has been strong evidence that the expectation hypothesis for interest rates does not hold for various time horizons, as for example documented in Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005), only bond market factors related to the risk of long term spread widening have been included in prior research (for example the term risk factor of Fama and French (1993))[8]. In this study, a factor relating to a shorter term premium will be included, which is directly related to the risk of shorter term interest rates rising unexpectedly. Another factor which has not been used in the

[7] The available offering of trading strategies by investment banks is often not disclosed. An exception is Barclays Capital which publishes their offering of trading strategies online: https://ecommerce.barcap.com/indices/index.dxml

[8] The expectation hypothesis refers to the idea that forward rates represent the expected future spot rates. If the expectation hypothesis holds, the expected return of holding a long maturity bond until maturity would be equal to the expected return of investing and rolling over shorter term bonds until the maturity of the longer term bond.

hedge fund literature to the best knowledge of the author, is related to the foreign exchange market where there has long been empirical evidence that there exists a forward premium as for example analyzed in Fama (1984). The forward premium can be extracted by following a carry trade strategy where an investor borrows in a low interest rate currency and invests the proceeds in a higher interest rate currency. The carry trade has been studied extensively in the academic literature and there is evidence that the carry trade premium is at least in part compensation for taking on specific systematic risk as shown in Christiansen, Ranaldo and Söderlind (2011) and Lustig, Roussanov and Verdelhan (2011). However, there are other strategies in the foreign exchange markets which have not gained much attention in the academic literature. The most prominent ones are trend following (momentum) and value strategies as described in the study of Pojarliev and Levich (2008), who examine the performance of currency funds. In previous research on hedge fund performance, the carry, momentum and value currency risk factors have not been included while often a factor related to the U.S. dollar index was used[9]. Commodities as an asset class have been included in hedge fund research, but the focus was mainly on a commodity market factor often represented by the Goldman Sachs Commodity Index (GSCI). As in the foreign exchange market, there is evidence that additional common risk factors exist when investing in commodities. The most prominent risk factors are related to momentum and backwardation strategies. Thereby, backwardation strategies in general go long commodities which are in backwardation, and go short those which are in contango. While it is shown in Gorton, Hayashi and Rouwenhorst (2007) that part of the excess return of commodity momentum and backwardation strategies is related to selecting commodities where inventories are low, Fuertes, Miffre and Rallis (2010) show that momentum and backwardation strategies are producing non-overlapping signals and can in fact significantly improve the risk return

[9] The U.S Dollar index represents a trade volume weighted basket of six currencies (EUR, JPY, GBP, CAD, SEK, CHF) quoted in USD.

characteristics of a portfolio when combined. Overall, in this study fifteen factors will be used whereby three currency, two commodities and one rates factors are newly introduced. How the various factors are defined is explained in section 1.2.2.

Proceeding of the Study

The new factors will be compared to an augmented Fung and Hsieh benchmark model by applying the model to single hedge fund data. Thereby, a best subset regression approach is used as explained in section 1.3.1 in order to mitigate the problem of over-fitting. As will be shown in section 1.4.1, the new factor model is superior compared to the benchmark model in terms of various metrics. Since there has been strong evidence in the existing literature that the exposure of hedge funds to risk factors changes over time, I will follow a similar approach as introduced by Bollen and Whaley (2009) and apply a discrete structural change model to hedge fund returns. In addition to the analysis done by Bollen and Whaley (2009), in this study hedge funds are allowed to change their exposure multiple times and the maximum number of factors allowed in the regression is not restricted. Following this approach, I am able to show that using a discrete structural change model on single hedge fund data identifies the same structural breaks as is the case in previous research[10]. Further, by applying the discrete structural change model on single hedge fund data and allowing for multiple changes in exposure, it is possible to track how the exposures to the various risk factors have changed over time. In addition, I will also investigate how the average alpha across hedge funds developed over time and thereby contribute to the discussion on diminishing hedge fund alpha[11]. In this respect I have found evidence that the number of funds generating significant positive alpha has

[10] For example in Fung et al. (2008), Meligkotsidou and Vrontos (2008) and Naik, Ramadorai and Stromqvist (2007).

[11] See for example Naik, Ramadorai and Stromqvist (2007), Fung et al. (2008), Zhong (2008) and Huber (2011).

decreased over time, whereby the findings differ with respect to the fund type and fund strategy under consideration.

In summary, the contributions of this study to the existing hedge fund literature are threefold. First, a new set of risk factors is being introduced which is shown to be superior compared to an augmented Fung and Hsieh factor model. Second, this study extends the discrete structural change model approach used by Bollen and Whaley (2009), by allowing for multiple breaks and allowing for more than three factors to be included in the multifactor model. Third, this study provides new evidence that the number of hedge funds generating positive alpha has decreased over time. The study will proceed as follows: In section 1.2 the used hedge fund data as well as the used risk factors will be explained in detail. Section 1.3 will continue by describing the applied econometric methodology and section 1.4 is providing the empirical results. Section 1.5 concludes by summarizing the results of this study.

1.2 Data

1.2.1 Hedge Fund Data Description

For this study I use the monthly hedge fund return data from the Center for International Securities and Derivatives Markets (CISDM) database. The sample spans the time period from January 1994 through March 2009. I only include funds which report their returns in USD since the used factor returns in this study are also USD based. All return series used in this study are based on arithmetic returns. Also, only funds which have at least 24 consecutive observations are included in the sample and funds which have missing months in their return history are dropped from the sample. This procedure ensures that we have enough data points for our estimation approach and that no bias is introduced from assumptions made in non-reported months. Further, following the approach of Huber (2011) I only include funds which have reported the value of assets under management and have at least once reported asset under management exceeding 5 million USD. This procedure ensures that very small funds with extreme performance are excluded from the sample. In addition, live and dead funds are considered which reduces the survivorship bias in the data set. One problem which arises when using the CISDM database is the fact that backfilling is not monitored. As explained in Fung and Hsieh (2000) this may lead to an instant history bias where funds start reporting their returns to a database after they have been successful and their performance history is ex post included in the database. Fung and Hsieh (2000) propose to drop the first 12 observations for each hedge fund in order to reduce the instant history bias. In this study I will not follow this approach since many funds already have relatively short reporting histories. Further, the study of Bollen and Whaley (2009) on hedge fund risk dynamics finds that not adjusting for the instant history bias does not significantly change their results. Another problem which arises in the analysis of hedge funds is the tendency for certain hedge funds to exhibit significant positive auto-correlation. Getmansky, Lo and Makarov (2004) propose a model to smooth returns and obtain smoothing adjusted Sharpe ratios. In this study, similar to Bollen and Whaley (2009), no return smoothing is applied to the hedge funds in the data sample since as noted by Bollen and Whaley (2009), the return smoothing has no qualitative effect on the results of their study.

Applying all the exclusionary criteria on the hedge fund data set results in a data sample of 4717 funds. Since part of this study is an extension of the work of Bollen and Whaley (2009) and to make results comparable to their research, I also use the fund type attributes available in the CISDM database to distinguish fund types. Namely these are Hedge Funds (HF), Commodity Trading Advisors (CTA), Fund of Funds (FoF) and Commodity Pool Operators (CPO). Since a further goal of this study is to analyze whether hedge fund alpha has decreased over time I will also differentiate funds according to their reported fund strategy in order to determine whether there are differences depending on the pursued strategies. The CISDM database distinguishes 28 different strategies[12]. In order to conduct a meaningful analysis, I only include fund strategies for which I have more than 100 funds in my sample after adjusting for the exclusionary criteria mentioned above. Further, for the purpose of brevity I will focus on strategies which are most likely prone to have exposure across asset classes in contrast to strategies which are focused on one asset class only. This leaves the following strategies which will be part of my analysis: Event Driven, Global Macro, Multi Strategy, Single Strategy and Systematic. Table 1.1 shows the summary statistics for the returns of the different fund categories under consideration based on the monthly excess return data. Excess returns are obtained by subtracting the one month treasury bill return from the returns in the database[13]. The table lists the equally weighted average of the mean, standard deviation, skewness, kurtosis and Sharpe ratio for each fund category under consideration. The last column lists the number of funds in the specific category which are included in the data sample. As seen in the table, the Single Strategy category has the highest Sharpe ratio followed by the HF and Event Driven categories. On the other side, the FoF category has the lowest Sharpe ratio on average. The kurtosis is the highest for the HF and FoF categories and the lowest for the Systematic category. The most negative skewness is found for the FoF category while the CTA, CPO and Systematic categories show the most positive skewness on average.

[12]See table 1.13 in the appendix for a list of strategies available in the CISDM database and the number of funds which are available after applying the exclusion criteria.

[13]The monthly treasury bill return data was obtained from the homepage of Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

Table 1.1: Hedge Fund Data Statistics

The statistics are the equally weighted averages for each considered hedge fund category of the mean monthly excess returns (mean), the standard deviation of the monthly excess returns (stdev), the skewness of the monthly excess returns (skew), the kurtosis (kurt) of the monthly excess returns and the Sharpe ratio (SR). The last column lists the number of funds in the specific category which are included in the data sample. The data sample covers the period from January 1994 to March 2009.

	mean	stdev	skew	kurt	SR	N
HF	0.0046	0.0408	−0.43	7.44	0.19	2699
CTA	0.0084	0.0547	0.28	5.19	0.14	442
FoF	0.0006	0.0243	−1.08	7.35	0.07	1386
CPO	0.0068	0.0499	0.29	5.16	0.13	190
Event Driven	0.0054	0.0332	−0.70	7.21	0.19	146
Global Macro	0.0062	0.0414	0.08	5.17	0.16	137
Multi Strategy	0.0023	0.0273	−0.90	7.31	0.12	1272
Single Strategy	0.0021	0.0316	−0.68	6.95	0.21	385
Systematic	0.0078	0.0521	0.29	4.80	0.15	350

1.2.2 Risk Factors Data Description

In this study, the benchmark model against which the new factors are compared is closely matching the factors used by Fung and Hsieh (2001) and Fung and Hsieh (2004). For the equity oriented factors I use the Standard & Poors 500 monthly total return index (ESP500) from Datastream and for the size spread factor (ESMB) I use the SMB factor as introduced in Fama and French (1993) as it is available on the homepage of Kenneth French[14]. The returns on the Standard & Poors 500 total return index were converted into excess returns by subtracting the one month treasury bill return as it is also available on Kenneth French homepage. For the bond oriented risk factor (BOND) I use the month-end to month-end change in the U.S. Federal Reserve 10-year constant-maturity yield and for the credit spread factor (CREDIT) I use the month-end to month-end change in the difference between Moody's Baa yield and the Federal Reserve's 10-year constant-maturity yield. The data used for both of these factors was downloaded from

[14] Link to Kenneth French homepage: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

the homepage of the board of governors of the federal reserve system[15]. The straddle based, or trend following risk factors as they are also referred to, are downloaded from the homepage of David Hsieh[16]. The bond, currency, commodity, rates and equity straddle factors are referred to in this study as STRBOND, STRCUR, STRCOM, STRRAT and STREQ respectively. The main difference to the factors used in Fung and Hsieh (2004) is the equity size spread factor where I use the SMB factor instead of the spread between the Wilshire Small Cap 1750 and the Wilshire Large Cap 750 index. Further, in this study the benchmark model is extended by the straddle factor on rates and equities, extending the Fung and Hsieh 7-factor benchmark model into an augmented 9-factor benchmark model. The summary statistics for the factors of the benchmark model are listed in table 1.2.

Table 1.2: Factor Statistics (Benchmark Model)

Listed are the mean, standard deviation (stdev), skewness (skew), kurtosis (kurt) and Sharpe ratio (SR) of the factors used in the Fung and Hsieh benchmark model over the time period from January 1994 to March 2009 based on monthly observations. ESP500 refers to the Standard & Poors 500 Total Return Index. ESMB is the small-minus-big equity premium factor introduced in Fama and French (1993). BOND is a fixed income related risk factor represented by the month-end to month-end change in the U.S. Federal Reserve 10-year constant-maturity yield. CREDIT is a credit risk factor calculated by taking the difference between Moody's Baa yield and the Federal Reserve's 10-year constant-maturity yield. STRBOND, STRCUR, STRCOM, STRRAT and STREQ are the option based straddle factors introduced in Fung and Hsieh (2001) related to bonds, currencies, commodities, rates and equities respectively.

	mean	stdev	skew	kurt	SR
ESP500	0.0025	0.0444	−0.77	4.14	0.06
ESMB	0.0004	0.0356	0.29	7.83	0.01
BOND	−0.0002	0.0024	−0.17	4.65	−0.07
CREDITSPREAD	0.0002	0.0018	3.01	22.99	0.11
STRBOND	−0.0097	0.1484	1.43	5.93	−0.07
STRCUR	0.0067	0.1980	1.37	5.69	0.03
STRCOM	−0.0003	0.1404	1.26	5.48	0.00
STRRAT	0.0341	0.2938	4.05	24.92	0.12
STREQ	−0.0475	0.1292	0.99	4.94	−0.37

[15] Link to the homepage of the board of governors of the federal reserve system: http:// www.federalreserve.gov/

[16] Link to the homepage of David Hsieh: http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm

The risk factors used in the new model introduced in this study are related to the four asset classes equities, fixed income, commodities and currencies. The description of how these factors are modeled are found below and the factor statistics can be found in table 1.3. For all risk factors monthly data for the time period January 1994 to March 2009 was used.

Equity Risk Factors

The equity related risk factors are the ESP500 and ESMB factors as they are also used in the benchmark model as well a value factor (EVMG) and a momentum factor (EMOM). For the value and the momentum factor I use the HML and Momentum factor respectively as they are available on Kenneth French's homepage.

Fixed Income Risk Factors

For the bond asset class factor RASSET, which represents a broad investment in low risk bonds, I use the US CITIGROUP USBIG CORPORATE AAA/AA 1-10Y: TOTAL RETURN Index adjusted by subtracting the risk free rate[17]. The risk factor Rates Long-term Spread (RLS), which represents the exposure to the longer horizon term premium, is created by taking the difference between the CITIGROUP USBIG TRSY. 10+Y - TOT RETURN Index minus the CITIGROUP USBIG TRSY. 1-3Y - TOT RETURN Index[18]. The risk factor representing the exposure to the short horizon term premium Rates Short-term Spread (RSS), is constructed by using monthly settlement price data on the 3-month Eurodollar Futures[19]. The idea is that there is a premium for holding a long position in a 3-month Eurodollar Future as a speculative long position is a bet that interest rates in the future will be lower than what is anticipated by the market. Holding a long position

[17] Datastream Code: USBC2A110.

[18] Datastream Codes: SBGT10P and SBGT13I.

[19] Datastream Code: IEDmmyy(PS) where mm and yy refer to the month a year of expiration of the contract.

in the 3-month Eurodollar Futures is therefore equivalent to taking on the risk of unexpected interest rate rises at the short end of the yield curve. The detailed construction of this risk factor is found in the appendix. The credit risk factor CCORP is used to represent the risk premium stemming from investments in higher risk corporate bonds, compared to investing in low risk government bonds. This risk factor is constructed by taking the difference between the CITIGROUP USBIG CORP. A 1-3 Y TOT RETURN Index minus the CITIGROUP USBIG TRSY- 1-3Y - TOT RETURN Index[20]. The last fixed income related risk factor is the high yield spread factor CHY, which accounts for the premium related to investing in high yield bonds compared to investing in investment grade corporate bonds. The risk factor is defined by taking the difference between the CITIGROUP HY MARKET : TOTAL RETURN Index and the CITIGROUP USBIG CORP. (LPF) - TOT RETURN Index[21].

Commodity Risk Factors

The first considered commodity risk factor COASSET reflects exposure to the overall commodity market and is represented by the Standard & Poors Goldman Sachs Commodity Index adjusted for the risk free rate[22]. In addition to the broad market risk factor, I introduce a commodity momentum factor (COMOM) and a commodity roll yield factor (COROLL) in this study. In defining the commodity momentum factor (COMOM) I follow closely the approach used by Miffre and Rallis (2007) in their study on momentum effects in the commodity markets. The detailed construction of this factor is described in the appendix. The used commodity data set consists of futures price data for 35 commodities[23].

The roll yield factor COROLL which I use in this study is related to the

[20]Datastream Codes: SBC1A13 and SBGT13I.

[21]Datastream Codes: USHYMKTRI and SBNLPFC.

[22]Datastream Code: GSCITOT.

[23]Table 1.14 in the appendix gives an overview of the included commodities and from which point in time they were included in the data set.

term structure strategy proposed in Fuertes, Miffre and Rallis (2010). The idea behind the strategy is to hold a long position in commodities which are in backwardation and to hold a short position in commodities which are in contango[24]. As Fuertes, Miffre and Rallis (2010) point out, one explanation for backwardation and contango shaped term structures is directly related to the inequality between the long and short positions of hedgers, which require the intervention of speculators to restore equilibrium[25]. As they put it:

> If commodity futures returns directly relate to the propensity of hedgers to be net long or net short, it becomes natural to design an active strategy that buys backwardated contracts and shorts contangoed contracts.

They propose a simple strategy which first determines at the end of each month which commodities are in backwardation and which ones are in contango. Then, they take equally weighted long positions in the top quintile of the most backwardated commodities and simultaneously take equally weighted short positions in the quintile of the most contangoed commodities. At the end of each month, the procedure is repeated and the positions are adjusted. The commodity roll yield factor used in this study is a variation of the approach of Fuertes, Miffre and Rallis (2010) and the detailed construction can be found in the appendix.

Currency Risk Factors

In this study, three currency risk factors are used to represent exposure to common strategies in the foreign exchange market often referred to as carry,

[24]Backwardation refers to a situation where the futures contracts with a longer time to maturity trade lower than the contract closest to expiration. Contango refers to the opposite situation, when the futures contracts with a long maturity trade at a higher price than the contract which is closest to expiration.

[25]Hedgers are market participants which have a direct business interest in the commodities they trade. An example is an oil exploration firm which needs to hedge the price of its future oil deliveries or an airline which wants to hedge its fuel expenses against sudden price increases. Speculators on the other side, do not have a direct interest in the underlying commodity and participate in the market merely to profit from price moves of the commodities.

momentum and value strategies. As Pojarliev and Levich (2008) show, these currency strategies are well able to explain the risk exposure of managed currency funds[26]. In addition, Kroencke, Schindler and Schrimpf (2011) show that including the three currency strategies in a portfolio of stocks and bonds can significantly improve the diversification of traditional portfolio allocations and that such portfolios are second and third order stochastic dominant when compared to portfolios not including these foreign exchange strategies. In this study, the currency strategies are implemented by using end-of-month spot and 1-month forward rate bid and ask prices from Datastream for 15 currencies. Table 1.15 in the appendix gives an overview of the used currency pairs. The detailed construction of the currency risk factors is described in the appendix. Table 1.3 gives an overview of all used factors in the new factor model together with their statistical properties for the time period from January 1994 to March 2009.

1.3 Methodology

In this section the methodologies used to conduct the factor model regression analysis as well as the structural change regression analysis are being explained. In section 1.4 the results of applying these methodologies on the data set will be presented.

1.3.1 Linear Multifactor Regression Model

In the existing literature on hedge fund performance appraisal, as referred to in section 1.1, linear multifactor models as introduced by Jensen (1968) are used to analyze the performance of hedge funds. In this study I will also use a linear factor model as described in equation 1.1 on page 17. Thereby, the excess return of the fund i is being regressed onto the excess return vector

[26] As a fourth factor they also introduce a foreign currency volatility risk factor.

Table 1.3: Factor Statistics (New Model)

Listed are the mean, standard deviation (stdev), skewness (skew), kurtosis (kurt) and Sharpe ratio (SR) of the factors used in the proposed new factor model over the time period from January 1994 to March 2009 based on monthly observations. ESP500 refers to the Standard & Poors 500 Total Return Index. ESMB, EVMG and EMOM refer to the small-minus-big equity, the high-minus-low equity as well as the momentum equity premium factors available on Kenneth French homepage. RASSET represents a broad investment in low risk bonds and is represented by the US CITIGROUP USBIG CORPORATE AAA/AA 1-10Y: TOTAL RETURN Index adjusted by subtracting the risk free rate. The long horizon term premium factor RLS, which represents the exposure to the longer horizon term premium, is calculated by taking the difference between the CITIGROUP USBIG TRSY. 10+Y - TOT RETURN Index minus the CITIGROUP USBIG TRSY. 1-3Y - TOT RETURN Index. The risk factor RSS, representing the exposure to the short horizon term premium, is constructed by using monthly settlement price data on the 3-month Eurodollar Futures. The credit risk factor CCORP, used to represent the risk premium stemming from investments in higher risk corporate bonds, is constructed by taking the difference between the CITIGROUP USBIG CORP. A 1-3 Y TOT RETURN Index minus the CITIGROUP USBIG TRSY- 1-3Y - TOT RETURN Index. The high yield spread factor CHY, which mirrors the premium related to investing in high yield bonds, is defined by taking the difference between the CITIGROUP HY MARKET : TOTAL RETURN Index and the CITIGROUP USBIG CORP. (LPF) - TOT RETURN Index. The COASSET risk factor, resembling an investment in the commodity asset class, is represented by the Standard & Poors Goldman Sachs Commodity Index adjusted for the risk free rate. COROLL represents the risk associated by investing in a roll yield strategy described by equation 1.5 in the appendix. COMOM represents a commodity momentum strategy similar to the one discussed in Miffre and Rallis (2007). CUR, represents a currency carry trade strategy where the trade signal is defined by equation 1.6 in the appendix. CURMOM is a currency momentum strategy where the trade signal is defined by equation 1.10 in the appendix. CURVAL represents a currency value strategy where the trade signal is defined by equation 1.11 in the appendix.

	mean	stdev	skew	kurt	SR
ESP500	0.0025	0.0444	−0.77	4.14	0.06
ESMB	0.0004	0.0356	0.29	7.83	0.01
EVMG	0.0024	0.0353	−0.03	5.69	0.07
EMOM	0.0080	0.0512	−0.57	7.45	0.16
RASSET	0.0016	0.0122	−0.28	8.42	0.13
RLS	0.0027	0.0241	0.17	6.44	0.11
RSS	0.0006	0.0038	−0.09	3.34	0.16
CCORP	0.0001	0.0082	−9.50	118.34	0.02
CHY	−0.0007	0.0247	−1.65	13.04	−0.03
COASSET	0.0020	0.0651	−0.46	4.29	0.03
COROLL	0.0041	0.0165	−0.39	6.33	0.25
COMOM	0.0097	0.0406	0.13	3.40	0.24
CUR	0.0018	0.0102	−0.46	4.56	0.17
CURMOM	0.0015	0.0133	1.26	9.97	0.11
CURVAL	0.0017	0.0141	0.51	4.49	0.12

of the risk factors $\mathbf{f}$.

$$r_{t,i} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_\mathrm{t} + \varepsilon_{t,i} \tag{1.1}$$

$$\varepsilon_{t,i} \sim N(0, \sigma_i^2) \tag{1.2}$$

The transposed vector $\boldsymbol{\beta}_i^T$ measures the constant exposure of a hedge fund to the risk factors in $\mathbf{f}_\mathrm{t}$ and α_i measures the performance of fund i which can not be attributed to the risk factors and is seen as the attribution to the management skill of a hedge fund manager. A significant positive α is therefore an indication that a hedge fund manager is able to add value and his/her performance is not just the result of having exposure to common risk factors.

An issue which arises when applying linear factor models with a higher number of factors to hedge fund returns is overparameterization, due to the fact that hedge fund return observations are only available on a monthly basis and for many hedge funds, track records only span a few years. The limitation in the degrees of freedom has been widely addressed in the existing literature on hedge funds. A common approach is to use regression models which select a subset of all available factors based on a variety of possible procedures. One of the more frequently used procedures is a forward stepwise regression approach as it is being applied for example in Liang (1999), Agarwal and Naik (2004), Titman and Tiu (2011) and Huber (2011). The mechanism of this approach adds and removes factors to the model based on the F-test or t-test significance. A typical procedure, for example applied in Agarwal and Naik (2004), would choose factors to be included based on the 5% t-test significance level. However, as it is pointed out in Hastie, Tibshirani and Friedman (2004), models which base the selection on the regression statistics are flawed, since they do not take into account multiple testing issues. An alternative consists of using an information criterion such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) to determine the inclusion of factors in the model. Bollen and Whaley (2009) follow this approach by choosing the best subset of all factors based on the smallest BIC value. Thereby, they limit the maximum number of factors to be included in the model to three. In this study, I follow the approach of Bollen and Whaley (2009) and use a best-subset approach using the Bayesian information criterion (BIC). In contrast to Bollen and Whaley (2009) however, I limit the number of factors to be included in the model to seven. As will be seen in the empirical results of section 1.4, for most hedge fund categories no more than 4 factors are included on average when using the BIC best-subset regression approach[27].

One additional issue pointed out by Hastie, Tibshirani and Friedman (2004) is that the standard inference of the linear regression model may be unreliable when conducting any subset regression approach as the search process for the optimal model is not being taken into account when calculating the inference metrics. This issue has not received much attention in the existing literature, however for example Agarwal and Naik (2004) explicitly address the issue. With regards to the statistical inference they state:

> The latter is a potential concern; however, it should only worsen the ability of the parsimoniously extracted factors to explain out-of-sample variation in hedge fund returns. Given that we obtain within-sample results that are consistent with other researchers and that we are able to replicate the out-of-sample performance of hedge funds, we believe that the benefits of using stepwise regression procedure outweigh its limitations.

In this context is was shown by Pötscher (1991) that when a consistent subset selection approach is used, the asymptotic distributions of the coefficients correspond to the asymptotic distributions which would apply when no subset selection was conducted[28]. As is described in Leeb and Pötscher (2005), the BIC best subset approach applied in this paper is consistent but the

[27] The subset selection process is conducted by using the leaps algorithm as introduced by Furnival and Wilson Jr (1974).

[28] A model selection procedure is regarded to be consistent when the procedure selects the true model with probability equal to 1 as $n \to \infty$.

finite-sample coefficient distributions for consistent selection procedures can be distorted. Whether the finite-sample coefficient distributions used for the inference measures are distorted depends on the probability of the selection approach to choose the true model. If this probability is high, the distortions are small and vice versa. Thus, in order to rely on the standard statistical inference in the regression model, we have to make the assumption that the BIC best subset selection model has a high probability of identifying the factors which are included in the true parsimonious model.

Being aware of the strong assumption needed in order to use the standard inference measures in the BIC best subset approach applied in this paper, I have constructed confidence intervals for α using a naive bootstrapping technique for the results presented in section 1.4.1. The results, which can be found in the appendix in table 1.18, indicate that the confidence bands for the intercept are wider compared to the case when the standard inference measures are used. However, the bootstrapped results confirm the finding that the proposed new risk factors are superior at explaining the performance of hedge funds compared to the benchmark risk factors.

1.3.2 Structural Change Regression Model

A number of papers have investigated whether the exposure of hedge funds to risk factors changes over time. One of the first approaches in this direction is found in Fung and Hsieh (2004), who use a variation of the standard test of cumulative recursive residuals to identify structural breaks in a broad hedge fund index. They identified two structural break points in the time period of January 1994 to December 2002. The first one is September 1998, which they associate with the failure of the LTCM hedge fund and the second one is March 2000, which they link to the end of the dot-com stock market bubble. Later research by Fung et al. (2008) confirm these breaks. Meligkotsidou and Vrontos (2008) compare different procedures for finding structural breaks on different hedge fund indices and their findings support the results from Fung and Hsieh (2004). Further, Huber (2011) uses a more recent data set to construct a hedge fund index and identifies two additional breaks. The first one in early 2004, which coincides with the beginning of the long equity bull market and the second one in August 2007, which he associates with a liquidity shock in the financial markets. Bollen and Whaley (2009) take a new approach at identifying structural breaks in hedge fund returns by investigating single hedge funds rather than hedge fund indices. They apply a changepoint regression approach as introduced by Andrews, Lee and Ploberger (1996). In their paper they compare the changepoint regression approach to a stochastic beta approach where the exposure of hedge funds to the risk factors is seen as an unobserved state variable following a stochastic process. They conclude that the changepoint regression approach is superior at identifying significant changes in the exposure of hedge funds to the risk factors. Bollen and Whaley (2009) do not explicitly identify the same structural breaks that were found in previous research. However, they only allow one break per fund which could lead to biased results when in reality more than one break occurs.

This study, besides introducing a new set of risk factors, intends to contribute further to the existing literature by expanding the work of Bollen and Whaley (2009). The goal is to analyze whether similar break points are found as in the existing literature when applying the new risk factors, whether the exposure to risk factors has changed over time and whether, when using the new factors, there is evidence that the alpha of hedge funds has declined over time. Thereby, I analyze single hedge funds as in Bollen and Whaley (2009), but I allow for up to three breaks per fund to occur. Instead of using the changepoint regression approach of Andrews, Lee and Ploberger (1996), which depends on F-tests to identify breaks, I apply the approach of Bai and Perron (2003), using the Bayesian information criterion to identify breaks. The multiple structural change model of Bai and Perron (2003) in the context of this study is applied as follows: First, the BIC best subset approach as described in subsection 1.3.1 is used on the entire data set of each hedge fund in order to identify the number of factors to be included in the model. The identified factors are then used in the structural change model. Before calculating the structural breaks using the approach of Bai and Perron (2003), the minimum segment size between break points and the maximum number of break points allowed for each fund need to be specified. In this study, I choose 12 months as the minimum segment size between break points and a maximum number of breakpoints P equal to 3. The minimum segment size limits the number of breaks for funds with less than 48 months of data. For example, a fund which only has 36 months of data can have a maximum number of $P = 2$ breakpoints. Having specified the parameters for the breakpoints, the dynamic programming approach of Bai and Perron (2003) then calculates in a computationally efficient manner for each possible number of p, the break dates which minimize the overall sum of squared residuals. In a final step, the optimal number of break points for a given fund is being found by comparing the optimal models with 0 to P break points to each other, based on the Bayesian information criterion (BIC). Thus for example, if the BIC value of the model with $p = 0$ is the lowest, no break is identified for this specific fund[29]. The multiple structural change regression model is described in equation 1.3. Thereby, Q refers to the optimal number of break points identified by comparing the possible

[29] The structural breakpoint approach of Bai and Perron (2003) is basically a two-step approach which works as follows: First, for all possible numbers of break points (in this study there are a maximum number of 3 breakpoints and the possible numbers of breakpoints are zero, one, two or three), the following calculation is being performed: Given the minimum segment size, it is being determined where the breakpoints need to be set, such that the overall sum of squared residuals are minimized. The overall sum of squared residuals are calculated by aggregating the sum of squared residuals of each segment which are determined by an ordinary least squares regression (OLS). In Bai and Perron (2003) an efficient algorithm is presented to find the minimum of the overall sum of squared residuals for a given number of breaks, however the goal could also be achieved by a grid search approach which calculates the sum of squared residuals for every possible variation of break points and then chooses those break points with the lowest overall sum of squared residuals. In the second step, the Bayesian information criteria (BIC) of the models (for each possible number of break point one model was determined, thus in this study four models were determined) found in the first step are compared, and the model with the lowest BIC value is the model chosen by the approach.

models with $p = 0, 1, ..., P$ breakpoints to each other based on the BIC metric. The index j refers to the sub-periods and α_j respectively $\boldsymbol{\beta}_j^T$ are the estimated coefficients in the sub-periods of the optimal model.

$$r_{t,i} = \alpha_{i,j} + \boldsymbol{\beta}_{i,j}^T \mathbf{f}_t + \varepsilon_{t,i} \tag{1.3}$$

$$\text{for } j = 1, .., Q+1$$

The two step approach of first determining the factors to be included in the model and then selecting the break points is the same as the one used by Bollen and Whaley (2009). In contrast to their study, I will re-estimate the BIC best subset model in each identified sub-period by allowing all factors to be considered for inclusion again. This approach has the disadvantage of not taking into account the possibility of a switching factor-set when determining the break-dates. Thus in effect, a possibly miss specified model is used to determine the break points. On the other side, the approach of Bollen and Whaley (2009) assumes that funds may only change their exposure to the initially included factors while not taking into account that over time, the factor-set to which hedge funds have exposure may change. For the purpose of this study it makes sense to allow the factor set to change, as the identified break points are similar to the ones found in previous studies. Further, allowing the factor-set to change will allow to gain insights on how the exposure to the risk factors of different hedge fund categories has evolved over time.

1.4 Empirical Results

The section on the empirical results is structured as follows. First, in subsection 1.4.1 the performance of the new risk factors at explaining the performance of hedge funds is compared to a benchmark model including the Fung and Hsieh straddle factors. Thereby, a static setting is assumed, where the exposure of the hedge funds to the different risk factors is not allowed to change over time. The used methodology in this part corresponds to the BIC best subset approach explained in subsection 1.3.1. Then, in subsection 1.4.2 the structural change regression model as described in subsection 1.3.2 will be applied on the data set. It will be shown that similar breakpoints are found when analyzing single hedge funds with the new factor model, as is the case in previous research. In subsection 1.4.3, it will be analyzed how the factor exposure of funds has changed over time and whether there are signs that the alpha of hedge funds has decreased over time.

1.4.1 Performance of the New Factor Model

In this subsection the new risk factors introduced in subsection 1.2.2 are compared to a nine factor benchmark model, which is based on the Fung and Hsieh (2004) model, in terms of their ability to describe the performance of single hedge funds. As described in subsection 1.2.1, I will distinguish four different fund types and five fund strategies in the analysis which results in a total of nine categories. The analysis in this subsection were derived by regressing the excess returns of each hedge fund in the data set onto the risk factors of the new model as well as the risk factors of the benchmark model. Thereby, the BIC best subset regression approach was used, which is explained in subsection 1.3.1. The results within each category were aggregated by calculating the equally weighted mean of the results in each category. The metrics by which the two models are compared to each other are the adjusted R-squared, the size of the alpha, the number of funds with an insignificant alpha at the 95% confidence level, as well as the number of factors which are included in the model. Table 1.4shows the aggregated results of the BIC best subset regression approach.

As can be seen from the numbers in table 1.4, the new risk factors are superior compared to the benchmark model in the given setting. The average adjusted R-squared is higher for each category when compared to the

Table 1.4: Summary Statistics of Static Regression

The table lists the adjusted R-squared, the average alpha, the percentage of funds with an insignificant alpha at the 95% confidence level (one-sided interval) as well as the number of factors which are included in the model for the new factor model and the benchmark factor model, each defined in subsection 1.2.2. The regressions are conducted by a best subset approach where the factors included in the model are selected by using the Bayesian Information Criterion (BIC). The data sample covers the period from January 1994 to March 2009. The confidence level is calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied.

	New Factor Model				Benchmark Factor Model			
	adj R2	alpha	insignificant alpha	factors	adj R2	alpha	insignificant alpha	factors
HF	0.40	0.0049	0.57	3.17	0.31	0.0067	0.48	1.96
CTA	0.24	0.0053	0.71	2.40	0.17	0.0086	0.55	1.60
FoF	0.52	0.0014	0.74	4.03	0.43	0.0035	0.47	2.80
CPO	0.25	0.0032	0.78	2.72	0.19	0.0070	0.57	1.95
Event Driven	0.44	0.0053	0.45	3.39	0.38	0.0067	0.38	2.28
Global Macro	0.28	0.0053	0.58	2.51	0.20	0.0072	0.53	1.68
Multi Strategy	0.48	0.0027	0.68	3.82	0.38	0.0050	0.42	2.59
Single Strategy	0.46	0.0027	0.67	3.59	0.35	0.0049	0.48	2.25
Systematic	0.26	0.0039	0.75	2.50	0.19	0.0080	0.53	1.71
Average	0.37	0.0039	0.66	3.13	0.29	0.0064	0.49	2.09

benchmark model. The same holds true for the average alpha which is lower for the new risk factors in each category when compared to the benchmark model.

Table 1.5: Difference in Alphas New Model Versus Benchmark Model - Paired T-Test

The table lists the mean difference of the alphas between the new model and the benchmark model for each category under consideration. A paired t-test is being conducted for each category, testing the hypothesis that the alphas resulting from the new model are not smaller compared to the alphas of the benchmark model. The resulting p-values are listed in the right column.

	mean difference of alphas	p-value
HF	−0.0018	0.00
CTA	−0.0033	0.00
FoF	−0.0021	0.00
CPO	−0.0038	0.00
Event Driven	−0.0014	0.00
Global Macro	−0.0019	0.00
Multi Strategy	−0.0023	0.00
Single Strategy	−0.0022	0.00
Systematic	−0.0041	0.00

The average alpha per month in the categories under consideration is 0.25% lower for the new factor model. Applying a paired t-test to the alphas, testing the hypothesis that the alphas stemming from the new factor model are not smaller than the alphas from the benchmark models corroborate these findings as shown in table 1.5. Also, the percentage of funds exhibiting an insignificant alpha at the 95% confidence level is higher for each category when compared to the benchmark. The new factor model does however include more factors in the model when compared to the benchmark model (3.13 vs. 2.09 factors).

When comparing the results of the analysis to the numbers found in previous research, the numbers for the old model are very similar (see for example Bollen and Whaley, 2009, p.1003). Similar to previous studies, the adjusted R-squared for the CTA and CPO categories are relatively low, while the adjusted R-squared for the FoF category is relatively high. The FoF category is also the category which includes the highest number of factors in the benchmark model as well as in the new factor model. In order to get a better insight into how the exposure to the risk factors varies across the different hedge fund categories in a static setting, tables 1.6 and 1.7 show the percentage of funds for each category with exposure to the factors and the average coefficient size for the new factor model and the benchmark factor model respectively.

As tables 1.6 and 1.7 show, the equity market factor ESP500 has been included for more than 50% of the funds except for the CTA, CPO, Global Macro and Systematic category. The same is true for the CCORP risk factor. Thereby, the ESP500 factor as well as the CCORP factor have a positive coefficient on average. Further, the momentum factor for commodities and currencies in the new factor model are included most frequently where you would expect them, namely the CTA, CPO and Systematic category which often implement trend following strategies. In the benchmark model it is equivalently the trend-following factors STRCUR and STRCOM which are included most frequently in the model for these categories. Further, it is interesting to note that the commodity asset class factor COASSET has been included in over 50% of the funds in the FoF and Multi Strategy categories for the period under consideration. The short term rates factor (RSS) does have some relevance for the Global Macro (20%) and CPO (25%) categories, while the currency value factor seems to have been of importance for the Event Driven category (25%). Interestingly, the newly introduced commodity roll risk factor (COROLL) is not included in more than 10% of the funds for any category.

In order to test the robustness of the results of this subsection, the calculations of this subsection were redone by using only data for the time period from January 1994 to June 2008. The reason for this approach was to take into account that the sharp equity market sell-off starting in September 2008 may distort the results, especially since our data sample ends in March 2009 which corresponds to the lows reached in global equity markets during the

Table 1.6: Percentage of Funds with Factor Exposure (New Model)

Panel A) shows for every factor in the new model the percentage of funds in each category for which it is included when the BIC best subset approach is applied in a static environment where the factor exposure is not allowed to change over time. Panel B) shows the average coefficient value of the included factors for each hedge fund category under consideration in the same setting as described for Panel A).

A) Percentage of hedge funds with factor exposure

	New Factor Model														
	ESP500	ESMB	EVMG	EMOM	RASSET	RLS	RSS	CCORP	CHY	COASSET	COROLL	COMOM	CUR	CURMOM	CURVAL
HF	0.56	0.28	0.24	0.24	0.17	0.14	0.15	0.33	0.19	0.27	0.07	0.09	0.13	0.13	0.18
CTA	0.19	0.09	0.05	0.17	0.13	0.10	0.13	0.11	0.14	0.25	0.10	0.26	0.16	0.33	0.17
FoF	0.76	0.24	0.24	0.44	0.17	0.14	0.12	0.64	0.19	0.58	0.04	0.11	0.06	0.10	0.19
CPO	0.17	0.15	0.06	0.21	0.15	0.09	0.25	0.08	0.17	0.23	0.07	0.27	0.17	0.48	0.16
Event Driven	0.67	0.34	0.15	0.08	0.18	0.17	0.17	0.36	0.37	0.27	0.07	0.12	0.08	0.12	0.25
Global Macro	0.36	0.16	0.14	0.19	0.15	0.12	0.20	0.15	0.14	0.23	0.09	0.09	0.18	0.13	0.18
Multi Strategy	0.68	0.22	0.22	0.40	0.17	0.14	0.14	0.56	0.18	0.56	0.05	0.12	0.08	0.11	0.18
Single Strategy	0.61	0.23	0.26	0.34	0.15	0.19	0.16	0.42	0.18	0.41	0.05	0.16	0.09	0.19	0.16
Systematic	0.19	0.08	0.05	0.16	0.14	0.10	0.14	0.10	0.16	0.26	0.09	0.28	0.18	0.37	0.18

B) Average coefficient values

	New Factor Model														
	ESP500	ESMB	EVMG	EMOM	RASSET	RLS	RSS	CCORP	CHY	COASSET	COROLL	COMOM	CUR	CURMOM	CURVAL
HF	0.51	0.32	−0.17	0.16	−0.02	0.04	0.45	2.14	0.31	0.19	0.16	0.10	0.72	−0.22	−0.08
CTA	0.11	0.26	0.24	0.35	0.73	−0.26	3.25	2.50	−0.43	0.25	−0.20	0.49	0.50	1.10	0.34
FoF	0.29	0.17	−0.17	0.16	−0.52	−0.20	1.15	1.15	0.10	0.10	0.05	0.12	0.22	−0.19	−0.08
CPO	0.17	0.47	0.25	0.28	−0.50	−0.20	6.11	2.88	−0.59	0.25	−0.63	0.33	0.65	1.24	0.49
Event Driven	0.37	0.34	0.11	0.03	0.78	0.16	0.11	1.63	0.60	0.15	0.12	0.04	−0.32	0.26	−0.29
Global Macro	0.29	0.27	−0.51	0.26	0.88	−0.15	0.22	1.65	−0.20	0.18	−0.71	0.29	0.50	0.22	0.81
Multi Strategy	0.29	0.15	−0.18	0.17	−0.25	−0.22	2.04	1.21	0.20	0.11	−0.12	0.15	0.50	0.13	−0.07
Single Strategy	0.34	0.22	−0.19	0.16	−0.31	−0.08	1.79	1.07	−0.19	0.13	−0.08	0.16	0.26	0.19	0.10
Systematic	0.00	0.32	0.24	0.43	1.27	−0.37	3.47	1.87	−0.46	0.25	−0.41	0.49	0.40	1.14	0.60

Table 1.7: Percentage of Funds with Factor Exposure (Benchmark Model)

Panel A) shows for every factor in the benchmark model the percentage of funds in each category for which it is included when the BIC best subset approach is applied in a static environment where the factor exposure is not allowed to change over time. Panel B) shows the average coefficient value of the included factors for each hedge fund category under consideration in the same setting as described for Panel A).

A) Percentage of hedge funds with factor exposure

	Old Factor Model								
	ESP500	ESMB	BOND	CRSPREAD	STRBOND	STRCUR	STRCOM	STRRAT	STREQ
HF	0.59	0.27	0.17	0.31	0.10	0.06	0.08	0.28	0.11
CTA	0.18	0.06	0.13	0.11	0.22	0.32	0.27	0.15	0.16
FoF	0.73	0.26	0.17	0.52	0.11	0.07	0.17	0.63	0.14
CPO	0.18	0.09	0.15	0.09	0.25	0.44	0.33	0.18	0.24
Event Driven	0.71	0.35	0.18	0.45	0.11	0.03	0.05	0.21	0.18
Global Macro	0.39	0.17	0.16	0.12	0.12	0.19	0.15	0.25	0.12
Multi Strategy	0.68	0.23	0.16	0.46	0.11	0.08	0.16	0.56	0.15
Single Strategy	0.55	0.23	0.15	0.37	0.09	0.10	0.18	0.43	0.14
Systematic	0.19	0.05	0.14	0.10	0.25	0.37	0.29	0.16	0.15

B) Average coefficient values

	Old Factor Model								
	ESP500	ESMB	BOND	CRSPREAD	STRBOND	STRCUR	STRCOM	STRRAT	STREQ
HF	0.48	0.30	−0.46	−6.38	−0.02	−0.01	0.06	−0.03	0.02
CTA	0.04	0.35	−0.27	−3.62	0.10	0.10	0.11	−0.04	−0.03
FoF	0.26	0.14	0.59	−3.94	−0.02	0.01	0.04	−0.02	0.04
CPO	0.09	0.52	−4.29	−3.58	0.10	0.09	0.10	−0.02	0.04
Event Driven	0.39	0.33	−0.71	−6.33	−0.09	−0.02	0.06	−0.04	0.01
Global Macro	0.30	0.16	−4.12	−3.38	0.02	0.05	0.11	−0.03	0.04
Multi Strategy	0.26	0.13	0.38	−4.17	−0.01	0.03	0.05	−0.02	0.04
Single Strategy	0.33	0.11	−0.94	−4.15	0.03	0.04	0.08	−0.02	0.03
Systematic	0.04	0.33	−1.77	−2.01	0.11	0.10	0.11	−0.03	0.03

sub-prime financial crisis. The summary statistics comparing the two models during the shortened time period can be found in the appendix in table 1.16. The results change in the sense that less funds have insignificant alphas when the observations from the second half of the year 2008 and the the first quarter of 2009 are excluded. However, the relative performance of the two models does not change in a meaningful way. Further, it may be argued that an enhanced benchmark model which uses additional factors such as a commodity market factor COASSET, the high yield factor CHY and the equity value factor EVMG will perform significantly better than the raw benchmark model consisting only of the factors proposed on David Hsieh's homepage. As it is shown in the appendix in table 1.17, enhancing the benchmark model with these factors, which have been used in previous research, does indeed improve its performance relative to the new factor model. However, using the straddle factors is still inferior to using the risk factors introduced in this study. Finally, as mentioned in subsection 1.3.1, the BIC best subset approach applied in this paper is consistent but the finite-sample coefficient distributions for consistent selection procedures can be distorted. In order to verify whether the confidence levels calculated in this paper by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors are robust, I also calculated the confidence intervals using the bootstrap percentile method as described for example in Efron (1981). As can be seen in table 1.18 of the appendix, the bootstrap results indicate that the number of funds with an insignificant alpha at the 95% confidence level are higher compared to the case when the standard inference measures are used. However, the bootstrapped results confirm the finding that the proposed new risk factors are superior at explaining the performance of hedge funds compared to the benchmark risk factors. Further, the average number of funds with an insignificant alpha at the 95% confidence level increases by 10% for the new factor model and 8% for the benchmark factor model, which indicates that the standard inference measures are only modestly underestimating the number of hedge funds with an insignificant alpha when compared to the bootstrapped inference measures. Therefore, in the subsequent analysis only the standard inference measures are reported.

1.4.2 Structural Change Regression Analysis

As analyzed in subsection 1.4.1, the set of new factors introduced in this study are on average superior at explaining the variation as well as the level of returns of the hedge fund categories under consideration. Therefore, in this subsection I will focus only on the new factor model when analyzing switching points in the exposure of single hedge funds to the risk factors. In subsection 1.3.2 it was pointed to the fact that many existing studies have mostly focused on identifying points in time where the exposure of hedge fund indices to risk factors change, thus not analyzing single hedge fund behavior. In this subsection I follow the approach of Bollen and Whaley (2009) and identify breakpoints for single hedge funds. Thereby, I extend the approach introduced by Bollen and Whaley (2009) by allowing three exposure changes instead of one, following the multiple structural change regression approach described in Bai and Perron (2003). Further, once the break points are identified, I allow the factor set to be included in each time period to change. The applied procedure is described in subsection 1.3.2.

Table 1.8: Number of Switches in Percent per Hedge Fund Category
The table lists the percentage of funds in each hedge fund category for which no switch, one switch, two switches and three switches in the factor exposure are found using the structural change regression approach of Bai and Perron (2003) described in subsection 1.3.2.

	No Switch	One Switch	Two Switches	Three Switches
HF	70.2%	22.4%	5.7%	1.7%
CTA	82.6%	16.3%	1.1%	0.0%
FoF	58.4%	33.2%	7.1%	1.4%
CPO	86.8%	9.5%	3.2%	0.5%
Event Driven	71.9%	24.7%	2.7%	0.7%
Global Macro	78.1%	14.6%	6.6%	0.7%
Multi Strategy	64.9%	26.5%	7.4%	1.3%
Single Strategy	72.5%	22.6%	4.2%	0.8%
Systematic	83.7%	15.1%	1.1%	0.0%

Table 1.8 gives an overview of the percentage of funds in each category for which zero, one , two or three switches have been identified using the multiple structural change regression approach. As shown, in the CTA, CPO and Systematic categories for over 80% of the funds no significant switching point was identified. The category with the highest percentage of switching funds is the FoF category followed by the Multi Strategy category. When looking at the hedge funds for which switches have been identified, it can be seen that for the majority only one switching point has been identified and less than 2% of all funds in any category exhibit three switches[30]. When comparing the results with those found in Bollen and Whaley (2009), it is worthwhile to note that they identified a higher percentage of funds to exhibit at least one switching point[31]. In order to examine whether the smaller percentage of switching funds is due to the chosen change point regression approach, I also conducted an Avg-F-Test as described in Andrews, Lee and Ploberger (1996) to identify whether there is at least one significant break point. The results for different significance levels are shown in table 1.9, whereby the numbers refer to the percentage of funds within each category for which a significant switch can be identified. As seen, the Avg-F-Test approach is on average more likely to identify significant break points even at the 99% probability level when compared to the BIC based approach followed in this study. Thus, the BIC based approach can be seen as conservative at identifying switching points compared to an Avg-F-Test approach.

[30]Table 1.21 in the appendix lists the average number of observations in the identified sub-periods for each hedge fund category.

[31]See Bollen and Whaley (2009) page 1011 for funds with significant switches. By using an F-test based procedure they found the following percentage of funds with significant switches at the 10% probability level: HF 43.3%, CTA 28.8%, FoF 49.9% and CPO 20.1%.

Table 1.9: Percentage of Switching Funds Using F-Test

The table lists the the percentage of funds for which one significant switching point is detected at the 90%, 95% and 99% confidence levels when using the Avg-F-Test approach described in Andrews, Lee and Ploberger (1996).

	90% conf. Level	95% conf. Level	99% conf. Level
HF	56.1%	47.6%	34.0%
CTA	50.0%	41.4%	28.5%
FoF	66.2%	57.4%	42.4%
CPO	42.1%	33.7%	22.1%
Event Driven	54.8%	44.5%	35.6%
Global Macro	55.5%	45.3%	27.7%
Multi Strategy	62.8%	55.0%	41.5%
Single Strategy	54.3%	43.9%	31.2%
Systematic	47.4%	38.6%	24.0%

Table 1.10 shows the performance metrics of the multiple structural change regression approach similar to the analysis in table 1.4 of subsection 1.4.1. For the multiple structural change regression analysis, two sort of results are reported. The first results, named arithmetic average for each fund, uses the simple average metrics of each identified sub-period for each fund. For example, if for a fund two breaks are identified, and the adjusted R-squares for each sub-period are 0.4, 0.3, and 0.6, then the used adjusted R-squared measure for this fund would equal to the arithmetic average of 0.43. The arithmetic average of the adjusted R-squared measure over all funds in a category then give the average adjusted R-squared measure shown in the table. This procedure is also applied to the alpha and insignificant alpha metrics[32]. For funds for which no break points were identified, the R-squared measures calculated by the BIC best subset approach described in subsection 1.4.1 are used. The second results, named time weighted average for each fund, calculates the average metric for each fund by weighing the results for each sub-period by the sub-period length. In this case, if the period lengths in the above example were 12, 18 and 34 months the used adjusted R-squared measure for this fund would be equal to $\frac{12}{64} \times 0.4 + \frac{18}{64} \times 0.3 + \frac{34}{64} \times 0.6 = 0.48$.

[32]For the insignificant alpha measure the described procedure is applied to the t-statistics of the intercept of each fund. Then, based on the average t-statistic, it is determined for each fund whether the value is significantly different from zero at the 95% confidence level (one-sided interval).

Table 1.10: Summary Statistics of Structural Change Regression

The table lists the average adjusted R-squared, the average alpha, and the percentage of funds with an insignificant alpha at the 95% confidence level (one-sided interval) when applying the structural change regression approach described in subsection 1.3.2 to the single hedge fund data in each hedge fund category using the new proposed risk factors. The arithmetic average section uses the arithmetic average of every metric for each fund. The time weighted average section uses the time weighted average metrics for each fund. The confidence level used to determine the significance of the alpha for each hedge fund is calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied.

	Arithmetic average for each fund				Time Weighted Average for each fund			
	adj R2	alpha	insignificant alpha	factors	adj R2	alpha	insignificant alpha	factors
HF	0.47	0.0049	0.56	3.53	0.44	0.0048	0.55	3.30
CTA	0.29	0.0084	0.70	2.71	0.27	0.0071	0.71	2.58
FoF	0.59	0.0000	0.75	4.26	0.54	0.0009	0.72	3.97
CPO	0.28	0.0034	0.75	2.85	0.26	0.0032	0.75	2.74
Event Driven	0.51	0.0062	0.47	3.60	0.47	0.0058	0.45	3.38
Global Macro	0.36	0.0053	0.58	3.06	0.33	0.0051	0.59	2.89
Multi Strategy	0.54	0.0017	0.70	4.03	0.50	0.0023	0.66	3.79
Single Strategy	0.52	0.0021	0.68	3.82	0.49	0.0023	0.66	3.62
Systematic	0.30	0.0061	0.75	2.81	0.29	0.0059	0.75	2.71

The results of table 1.4.2 show that the adjusted R-squared measures are higher for the arithmetic as well as the time weighted average approach when compared to the results of the undynamic model of subsection 1.4.1. However, interestingly, the average alpha is generally lower and the number of funds with an insignificant alpha is higher when using the undynamic model. The exception are the FoF, Single Strategy and Multi Strategy categories where the average alphas of the dynamic model is lower when compared to the undynamic model. Thus, the analysis found evidence that the dynamic approach is able to explain more of the variation in hedge fund returns while there does not seem to be an advantage of the dynamic modeling approach at explaining the level of hedge fund returns.

In their study, Bollen and Whaley (2009) found some evidence that the performance of funds for which a significant switch point is identified is superior when compared to funds with no switch point. Thereby, they measure the performance by the Sharpe ratio. As is shown in table 1.11 the result is ambiguous for the approach used in this paper. When looking at the four fund type categories also analyzed in Bollen and Whaley (2009), there is some indication that for the HF, CTA and FoF categories, switching funds do have a superior Sharpe ratio when compared to the non-switching funds. However, when analyzing the results for the categories introduced in this study, the result is mixed. Also, it needs to be pointed out that there is a possible bias in this analysis as more successful hedge funds are likely to stay in business for a longer time and therefore are more likely to change their exposure at some point in time[33].

[33]Table 1.22 in the appendix gives an overview of the average performance of funds during the identified sub-periods.

Table 1.11: Performance of Switchers and Non-Switchers

The table shows the average performance statistics of hedge funds for which no switch is identified and for hedge funds for which switches are identified when using the structural regression analysis described in subsection 1.3.2. The statistics are the equally weighted averages for each considered hedge fund category of the mean monthly return (mean), the standard deviation of the monthly returns (stdev), the skewness of the monthly returns (skew), the kurtosis (kurt) of the monthly returns and the Sharpe ratio. The last column lists the number of funds in the specific category which are included in the data sample. The data sample covers the period from January 1994 to March 2009.

	Non-Switchers						Switchers					
	mean	stdev	skew	kurt	SR	N	mean	stdev	skew	kurt	SR	N
HF	0.0045	0.0403	−0.31	5.63	0.18	1894	0.0048	0.0419	−0.71	11.71	0.22	805
CTA	0.0079	0.0551	0.22	4.64	0.13	365	0.0104	0.0528	0.58	7.80	0.22	77
FoF	0.0005	0.0244	−0.79	5.48	0.07	809	0.0007	0.0241	−1.49	9.97	0.07	577
CPO	0.0069	0.0500	0.23	4.49	0.13	165	0.0056	0.0494	0.68	9.55	0.12	25
Event Driven	0.0048	0.0326	−0.51	5.50	0.18	105	0.0070	0.0347	−1.17	11.59	0.22	41
Global Macro	0.0060	0.0412	0.00	4.91	0.16	107	0.0068	0.0419	0.37	6.11	0.15	30
Multi Strategy	0.0026	0.0282	−0.61	5.55	0.12	825	0.0017	0.0257	−1.44	10.56	0.11	447
Single Strategy	0.0023	0.0334	−0.57	5.52	0.24	279	0.0014	0.0269	−0.97	10.71	0.13	106
Systematic	0.0076	0.0521	0.27	4.33	0.14	293	0.0088	0.0522	0.43	7.23	0.16	57

In the last part of this subsection, the switching frequency over the time period of January 1994 until March 2009 in the different hedge fund categories will be examined over time. In the estimation setup there need to be at least 12 observations in each identified sub-period. Thus, the first date on which a switch can occur is January 1995 and the last date is March 2008. As was pointed out in subsection 1.3.2, Fung et al. (2008), Meligkotsidou and Vrontos (2008) and Huber (2011) identified similar breakpoints where hedge fund indices change their exposure to risk factors. Common break points found were located in the fall of 1998 (LTCM collapse) and the spring of 2000 (dot-com bubble collapse). Huber (2011) also identified additional breaks at the beginning of 2004 (Start of equity bull market) and the summer of 2007 (Liquidity shock). The results of this study with regard to the switching frequency of individual hedge funds are shown in figure 1.1. The bar plots show for each category the number of funds for which a break was identified at a given month as a percentage of all funds in the category for which a switch is feasible in that month. In the HF category it is seen that the switching frequencies are remarkably higher in the periods of fall 1998 as well as in spring/winter 2000. In addition, there is a large increase in the switching frequency starting in the summer of 2007, and reaching the maximum on the final switching date, April 2008. Thus, the results are in-line with the findings of previous research based on index level analysis. The results for the FoF, Multi Strategy and Single Strategy categories confirm the results of the HF category by showing increased switching frequencies around the same points in time. For the Event Driven and Global Macro categories the results are similar, however, due to the lower number of funds in these samples, the results are less meaningful. Interestingly, for the CTA, CPO and Systematic categories, the results differ to some extent. Those categories show high switching activity at the very beginning of the period which may be an artifact of the small sample size available for that time period. However, looking at the switching frequencies over time, there seems to be increased activity in the fall of 1998 while the dot-com bubble burst of 2000 as well as the sub-prime crisis starting in 2007 do not seem to have triggered similar switching activity as was the case for the other categories. Thus, not only do CTA, CPO and Systematic hedge funds switch their exposure to the risk factors less frequently compared to the other categories under consideration, but they also do not occur at the same points in time.

In order to validate the qualitative results found by analyzing the distributions of the switching frequency in figure 1.1, I also conducted a regression analysis where the found percentage number of switches at each date are regressed onto the level of the VIX index as well as the three-month TED spread. The VIX index measures the near term expected volatility of the broad equity market by extracting and combining the implied volatility of short term options on the S&P 500 index[34]. The VIX index is seen by many market participants as a fear gauge in the sense that the level of the index rises significantly whenever the uncertainty and distress in the financial markets increase. The three-month TED spread is defined as the difference between the three-month USD LIBOR rate and the three-month US treasury bill interest rate. The TED spread is also perceived as a good indicator of stress in the banking and financial system in general, as it reflects the perceived credit risk of lending money to commercial banks. In figure 1.2 in can be seen that the VIX index as well as the TED spread tend to spike around the market events discussed above. The regression analysis, for which the results are found in table 1.12, provide further evidence confirming the prior results that hedge funds tend to switch the exposure to risk factors during times of big market shifts. For the HF and and Systematic categories, the VIX index level as well as the TED spread are significant at the 95% confidence level for explaining the percentage of switching funds. For the CTA, FoF, Multi Strategy as well as the Systematic category the TED spread is significant at the 95% confidence level, while the VIX index level is not.

[34]More information on the construction of the VIX index can be found under http://www.cboe.com /micro/VIX/vixintro.aspx .

For the Event Driven, Global Macro as well as the CPO categories, neither the VIX index level nor the TED spread has been found to be significant at explaining the percentage of funds which switch the exposure to the risk factors at a given month in time. When looking at the adjusted R-squared measures of the regressions, it can be seen that only for the HF, FoF, Multi Strategy and Single Strategy categories a meaningful amount of the variation in the switching frequency can be explained by the two factors. For all other categories the adjusted R-squared is found to be close to 0.

When comparing the results of this study with regards to the switching frequency of single hedge funds to the results of Bollen and Whaley (2009), they differ substantially. Bollen and Whaley (2009) find that changes to the risk factor exposures do not generally occur at the dates identified in previous research focused at hedge fund indices. They come the the conclusion that

> ... it appears that individual managers are actively changing factor loadings, as opposed to factor loadings shifting as a consequence of breaks in the time-series of underlying strategy returns.

The findings of this study suggests that for the majority of hedge fund categories, the switching frequency of single hedge funds is significantly elevated around the breakpoints identified in the previous literature and that therefore, these switches may be driven by structural breaks in the returns of the underlying strategies related to global market events. With regards to the CTA, Systematic and CPO categories, the results of this study are in-line with the finding of Bollen and Whaley (2009). For these categories the switching frequencies tend to be lower and more evenly spread out over time. A question which needs to be addressed is the underlying driver for the differing results of this study with regards to the HF and FoF category when compared to the study of Bollen and Whaley (2009). As it is shown in detail in the appendix, there is strong evidence that the differences in the findings are mainly driven by the choice of the factor set in the factor model as well as the number of break points allowed in the estimation procedure.

Figure 1.1: Switching Frequency of Hedge Funds Over Time

The bar plots show for each hedge fund category the percentage of hedge funds at each point in the data sample for which a switch in factor exposure is identified using the structural change regression approach described in subsection 1.3.2. The percentage measure is defined by the number of switching funds divided by the number of funds in the sample for which switching the exposure was feasible at the given month.

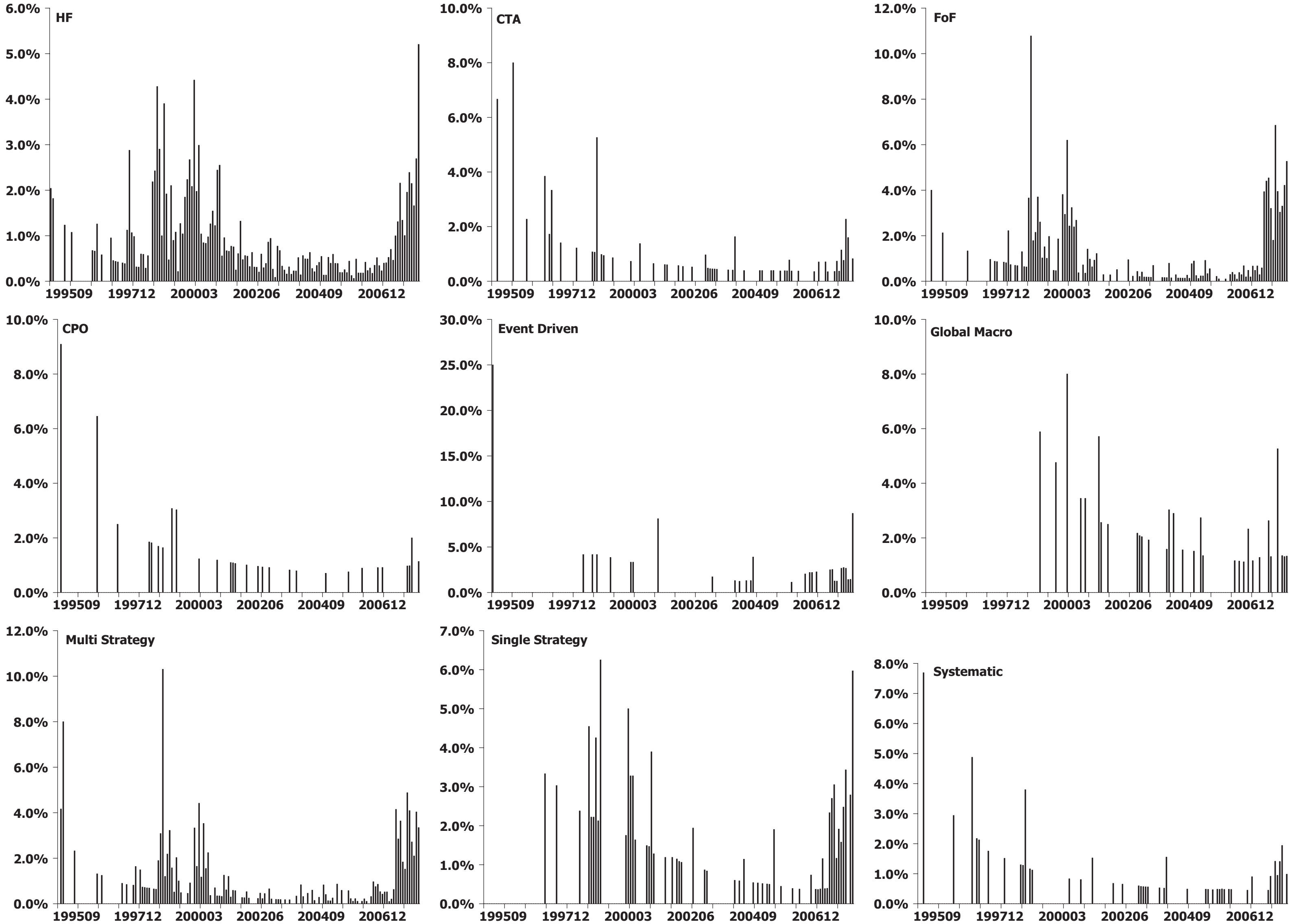


Figure 1.2: VIX Index and TED Spread

The bar plots correspond to the one for the HF category in figure 1.4.2. The graph on the left shows the VIX index for the time period from January 1995 to April 2008, whereby the level of the index is shown on the right hand axes of the figure. The graph on the right shows the TED spread for the time period from January 1995 to April 2008, whereby the level of the index is shown on the right hand axes of the figure. The data for the VIX index as well as the TED spread are from Bloomberg.

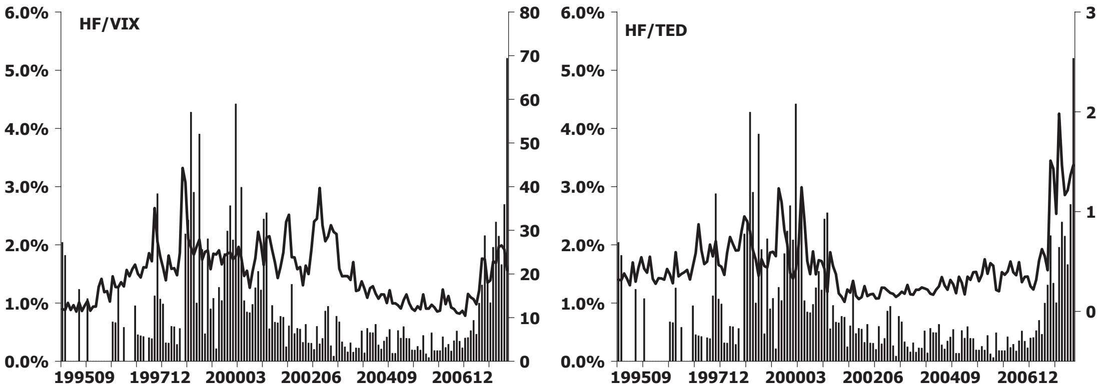


Table 1.12: Regression of Switching Frequencies on VIX/TED

This table shows the results of an ordinary least squares regression of the percentage of exposure switching funds in each category onto the VIX index, the TED spread and a constant (intercept) for each hedge fund category. The p-values (two-sided interval) are calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied.

		intercept	VIX	TED	adj R2
HF	coef	−0.0057	0.0004	0.0132	0.32
	p-val	0.0124	0.0017	0.0000	
CTA	coef	0.0029	0.0000	0.0042	0.00
	p-val	0.3029	0.7077	0.0099	
FoF	coef	−0.0088	0.0003	0.0275	0.35
	p-val	0.0054	0.0915	0.0000	
CPO	coef	0.0031	0.0000	0.0020	−0.01
	p-val	0.3797	0.7814	0.1847	
Event Driven	coef	0.0040	−0.0001	0.0112	0.01
	p-val	0.6365	0.7014	0.0559	
Global Macro	coef	0.0004	0.0001	0.0050	0.01
	p-val	0.8833	0.3642	0.2384	
Multi Strategy	coef	−0.0044	0.0001	0.0213	0.23
	p-val	0.1831	0.2917	0.0000	
Single Strategy	coef	−0.0058	0.0003	0.0139	0.16
	p-val	0.0322	0.0398	0.0006	
Systematic	coef	0.0013	0.0000	0.0040	0.01
	p-val	0.5916	0.8453	0.0075	

1.4.3 Factor Exposure and Performance Appraisal Over Time

The dynamic modeling approach applied in this study takes into account that hedge funds may change their investment strategies and that therefore, their exposure to common risk factors may shift over time. This dynamic approach does not only allow to investigate at what points in time hedge funds change their exposure, as discussed in subsection 1.4.2, but it can also give insights into how the average alpha of hedge funds has changed over time. These insights are linked directly to the discussion in the existing literature on whether the abnormal performance of hedge funds has decreased over time. In this context, the study of Fung et al. (2008) analyzed a data set of fund of hedge funds for the time period of 1995 to 2004 and found that the average fund was only able to deliver a positive alpha in the time period between October 1998 and March 2000. The study of Naik, Ramadorai and Stromqvist (2007) points into the same direction, as they show that the alpha of hedge fund strategy indices has decreased for the time period of 2000 until the end of 2004. Zhong (2008) uses single hedge fund data and confirms the result of decreasing alpha for the time period of January 1994 to December 2005. He links his findings to capacity constraints in the market. In contrast, Huber (2011) uses a more recent data set spanning the period from January 1994 to September 2008 and does not find further evidence of decreasing alpha for the period after 2004 for single hedge funds[35]. This study will contribute to the discussion by giving insights on how the average alpha of single hedge funds in the different categories have evolved over time, when allowing for multiple structural breaks in the return time series.

The following analysis will show the evolvement of the average adjusted R-squared as well as the percentage of funds with an insignificant positive alpha at the 95% confidence level for the time period of January 1994 to March 2009 for each category. Thereby, also the number of active funds which are used to calculate the two metrics at each point in time are tracked[36]. The adjusted R-squared values are derived from the calculations in section 1.4.2. Accordingly, a fund for which no significant switch was detected contributes the same metric to each month during which it is active. If a switch is detected, the fund contributes the new metrics from the break point onwards. Thus, the average adjusted R-squared changes over time by new funds becoming active, funds becoming inactive and funds contributing new adjusted R-squared measures due to identified breaks in their return series. The same procedure is used to calculate the percentage of funds with an insignificant

[35] Using a rolling stepwise regression approach.

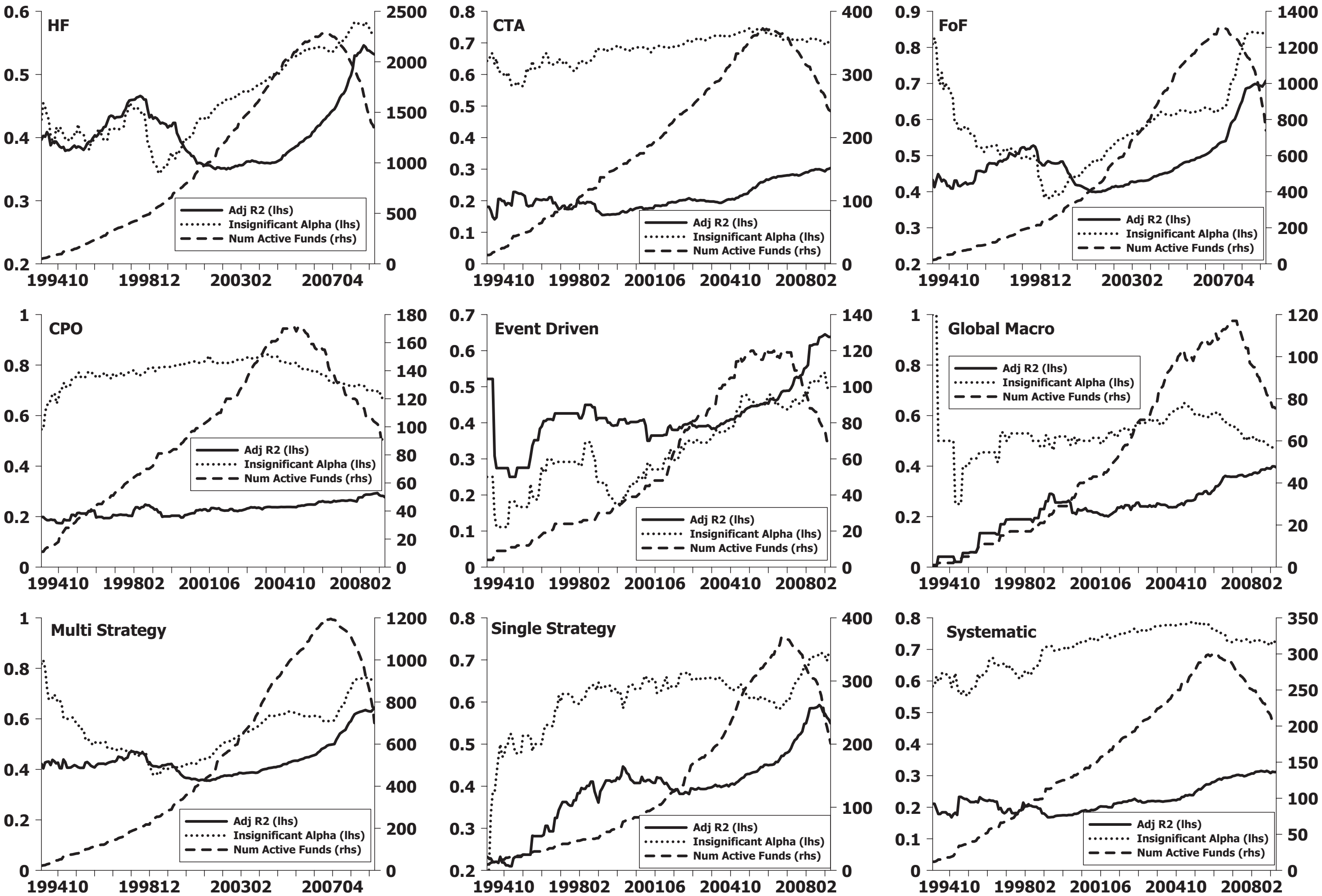
[36] As is seen in the graphs in figure 1.3, the number of active funds in the data set starts shrinking in all categories from 2007 onwards. The reason for this is twofold: First, the total number of hedge funds started to decline due to an increased number of liquidations during the sub-prime crisis and second, only funds with at least 24 monthly observations are included in the data set. Therefore, any fund which started reporting after April 2007 is not being included in the data set.

alpha in each category at each point in time. The graphs in figure 1.3 show the evolvement of the adjusted R-squared, the number of funds with an insignificant alpha as well as the number of active funds over the time period from January 1994 to March 2009.

As can be seen in the graphs, the HF, FoF, Event Driven and Multi Strategy categories show a similar pattern in the evolvement of the percentage of funds with an insignificant alpha. In each category, there is a drop in the number of funds with an insignificant alpha starting around October 1998 and lasting until the beginning of the year 2000. These findings confirm the findings of Fung et al. (2008) who found that in their data sample of fund of hedge funds, the most significant positive alpha was observed during the same period. Similarly, in all of these categories the percentage of funds for which an insignificant alpha was estimated increases from the year 2000 onwards, reaching a peak near the end of the data series in 2009. Thereby, it is worthwhile to point out that for the HF and the Event Driven category the increase is continuous, while for the FoF and Multi Strategy categories, there seems to be a period from around June 2004 until mid 2007 when the decrease in significant alpha funds does not continue but even slightly declines for the Multi Strategy category. Thus, while there is evidence that the average alpha for the HF, FoF, Event Driven and Multi Strategy categories has decreased for the time period after the year 2000, I also found some evidence that support the findings of Huber (2011), which indicate that at least for the FoF/Multi Strategy categories the average alpha has not further decreased during the period from mid 2004 until mid 2007. Looking into the evolvement of the average adjusted R-squared measure of the HF, FoF, Event Driven and Multi Strategy categories, there also seem to be similar patterns. For these categories the average adjusted R-squared values increase in the run-up to the LTCM crisis in September 1998 but fall afterward, until the end of 2001. Thereafter, they increase again, reaching a high during the peak of the sub-prime crisis at the end of the year 2008. Thus, for these

Figure 1.3: Performance Appraisal Over Time

The following plots show the average adjusted R-squared, the percentage of funds with an insignificant alpha at the 95% confidence level as well as the number of active funds in the data sample for each hedge fund category at each point in time. The average adjusted R-squared measure as well as the percentage of funds with an insignificant alpha at each point in time are derived from the structural change regression approach described in subsection 1.3.2. The analyzed time period is January 1994 to March 2009.



categories, the power of the chosen risk factors at explaining the variability of returns has increased over time.

The CTA, CPO and Systematic categories show different patterns when compared to the categories discussed above. For all three categories the percentage of funds with an insignificant alpha is relatively high at over 60% for almost the entire period. Thereby, the values show a slight increase over the entire period, however this increase is small compared to the HF, FoF, Event Driven and Multi Strategy categories. The average adjusted R-squared measures also show a very similar pattern for the CTA/Systematic categories: In the period from January 1994 to December 1998, the measure is alternating between increasing and decreasing. After December 1998, the average adjusted R-squared increases until the end of the data set in March 2009. The average adjusted R-squared measure for the CTA category increases from 0.16 in December 1998 to 0.31 in March 2009. In comparison to the HF and FoF categories, the number of funds with insignificant alphas has not moved much over the observed period, while the average adjusted R-squared measure has increased substantially over the period after December 1998.

The Global Macro and Single Strategy categories show somewhat different patterns in the movement of the measures under consideration. Thereby, it needs to be pointed out that the number of observations at the beginning of the time period are scarce for the two categories. The Global Macro category shows an increase of the average adjusted R-squared measure over time, while interestingly, the number of funds with insignificant positive alpha decreases during the sub-prime crisis of 2008-2009. This result is in-line with the findings of Sandvik et al. (2011), who find that Global Macro funds have performed particularly well in bear market environments. For the Single Strategy category, the percentage of funds with insignificant alpha increases until January 2003 and then declines slightly until January 2007. The measure then shows a sharp increase related to the sub-prime crisis. The average adjusted R-squared measure for the Single Strategy category increases until

January 2000, from where the measure declines until December 2002. Beginning in January 2003 the measure then increases to peak in October 2008, which is similar to the development as it is observed in the HF, FoF, Event Driven and Multi Strategy categories.

In summary, the results of the approach taken in this study are coherent with the findings of previous research. Namely, evidence is found that the average alpha has decreases over time for a majority of hedge fund categories under consideration. Also, the findings of Fung et al. (2008), showing that fund of hedge funds performed better during the period from October 1998 until March 2000, can be confirmed by the results of the approach taken in this study. The results of Huber (2011) could not be fully confirmed, as I do find that the percentage of funds generating significant positive alpha has decreased over time. However, as pointed out, for the FoF and Multi Strategy categories it seems that the notion of non-decreasing alpha seems to hold for the period between 2004 and mid 2007. This study also illustrates that different hedge fund categories show different developments over time with regards to the percentage of funds which produce significant alpha. The CTA, CPO and Systematic categories, while generally having high percentage of funds with insignificant alphas over the entire time period, did not show deteriorating performance in the sub-prime crisis of 2008/2009 as most other categories did. The Global Macro category even showed enhanced performance during the most recent crisis. In order to test the robustness of the results, the same calculations were conducted on a data set which only uses data until June 2008, therefore dropping the observations from July 2008 to March 2009. The idea behind dropping these observations is motivated by the fact that equity markets dropped by almost 38% (S&P 500) during this time period and that the observed volatility in the equity market was the highest on record since the Great Depression. The results of the constrained data set are found in the appendix in figure 1.13. Generally the results described so far in this subsection still hold. The average adjusted R-squared measure generally increased after the year 2001. Also, the percentage of funds with an insignificant alpha increased over the same period in the HF, FoF, Event Driven and Multi Strategy categories. However, the sharp increase in the percentage of insignificant alpha funds for these categories towards the end of the sample is not present when using the constrained data set. For the HF and Multi Strategy categories, there is even a decline in the percentage of insignificant alpha funds towards the end of the sample. This indicates that the escalation of the sub-prime crisis in September 2008 did have a significant impact on the performance of hedge funds which is also reflected by the high number of switches which are detected during this period (see figure 1.4.2).

The approach of tracking the average adjusted R-squared and the percentage of funds with insignificant positive alpha over time can be expanded to tracking the exposure to the risk factors over time. The procedure is similar to the one described above. For each month in the data set, it is being determined what percentage of funds have positive exposure to a certain factor. The same is done with regards to the negative exposure to risk factors. The resulting time series reveal insights into how the importance of the different risk factors in the model have changed over time. In order to increase the readability, the results for each hedge fund category are split between equity related factors, fixed income related factors, commodity related factors and currency related factors. Thereby, the lines above the zero mark are representing the percentage of funds with positive exposure while the lines below the zero mark show the percentage of funds with negative exposure over time. The results for the four fund type categories are shown in figure 1.4 and 1.5 respectively. The results for the five fund strategy categories are shown in figures 1.6, 1.7 and 1.8 respectively.

Figure 1.4: Factor Exposure Over Time (by Fund Type - HF and CTA)

The following plots show the percentage of hedge funds with positive and negative exposure to the new risk factors proposed in this study for the HF and CTA fund types over time. The four plots per hedge fund type show the exposure to the equity-, fixed income-, commodity- and currency-related factors separately. The lines above the zero-mark display the percentage of funds with positive exposure and the lines below the zero-mark display the percentage of funds with negative exposure. The percentage exposure to the risk factors are derived from the structural change regression approach described in subsection 1.3.2.

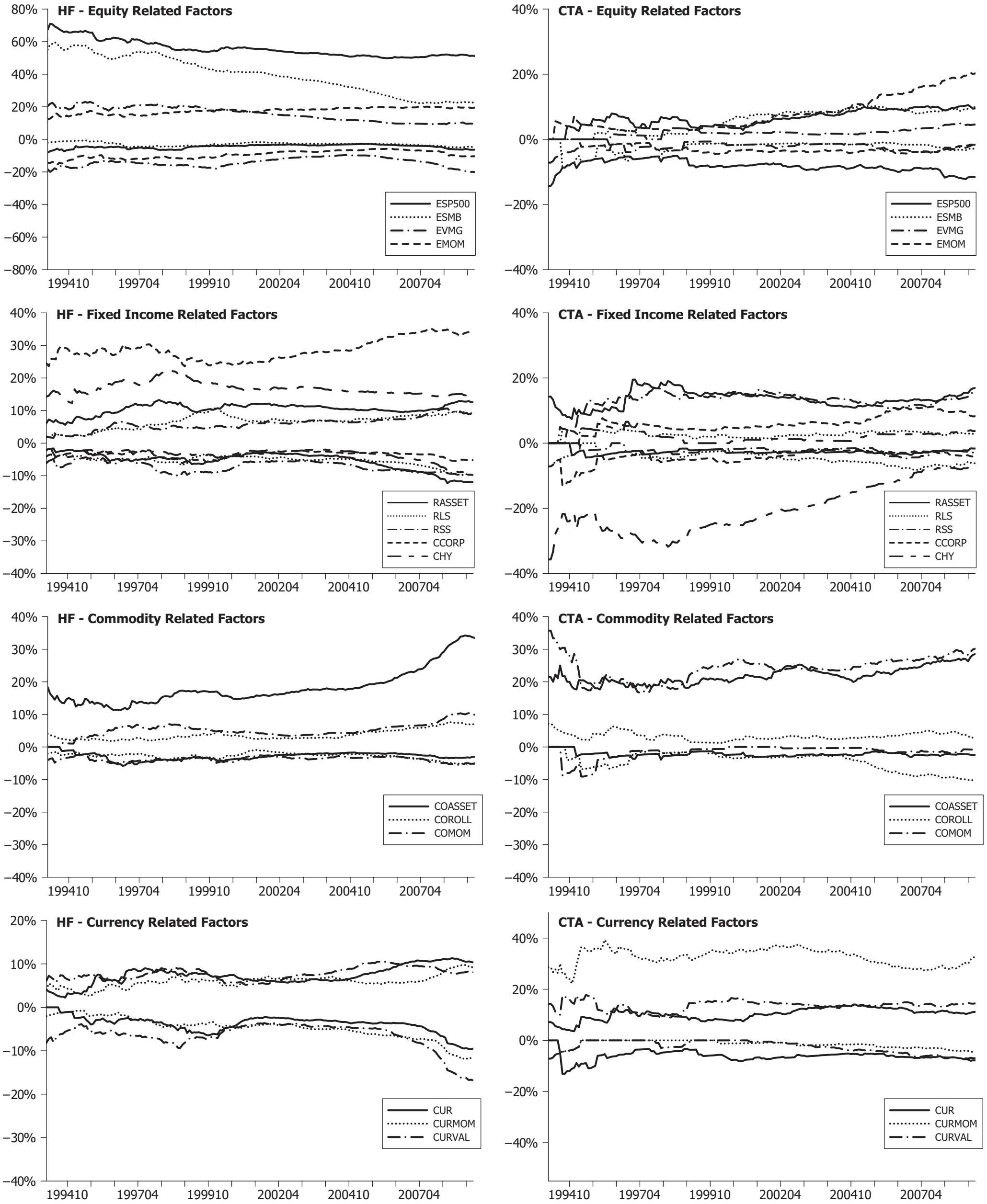


Figure 1.5: Factor Exposure Over Time (by Fund Type - FoF and CPO)

The following plots show the percentage of hedge funds with positive and negative exposure to the new risk factors proposed in this study for the FoF and CPO fund types over time. The four plots per hedge fund type show the exposure to the equity-, fixed income-, commodity- and currency-related factors separately. The lines above the zero-mark display the percentage of funds with positive exposure and the lines below the zero-mark display the percentage of funds with negative exposure. The percentage exposure to the risk factors are derived from the structural change regression approach described in subsection 1.3.2.

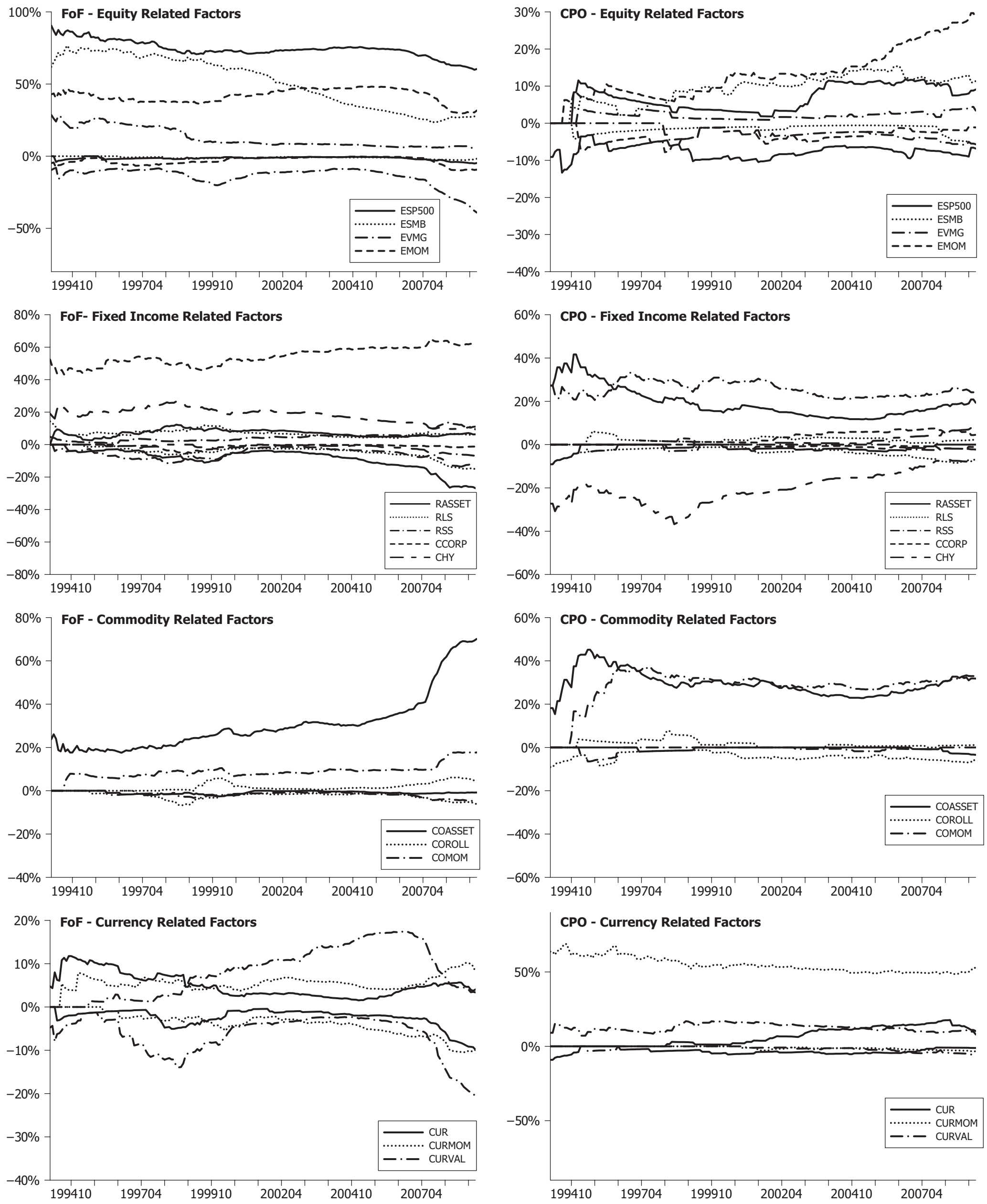


Figure 1.6: Factor Exposure Over Time (by Strategy Type - Event Driven and Global Macro)

The following plots show the percentage of hedge funds with positive and negative exposure to the new risk factors proposed in this study for the Event Driven and Global Macro strategy types over time. The four plots per hedge fund strategy type show the exposure to the equity-, fixed income-, commodity- and currency-related factors separately. The lines above the zero-mark display the percentage of funds with positive exposure and the lines below the zero-mark display the percentage of funds with negative exposure. The percentage exposure to the risk factors are derived from the structural change regression approach described in subsection 1.3.2.

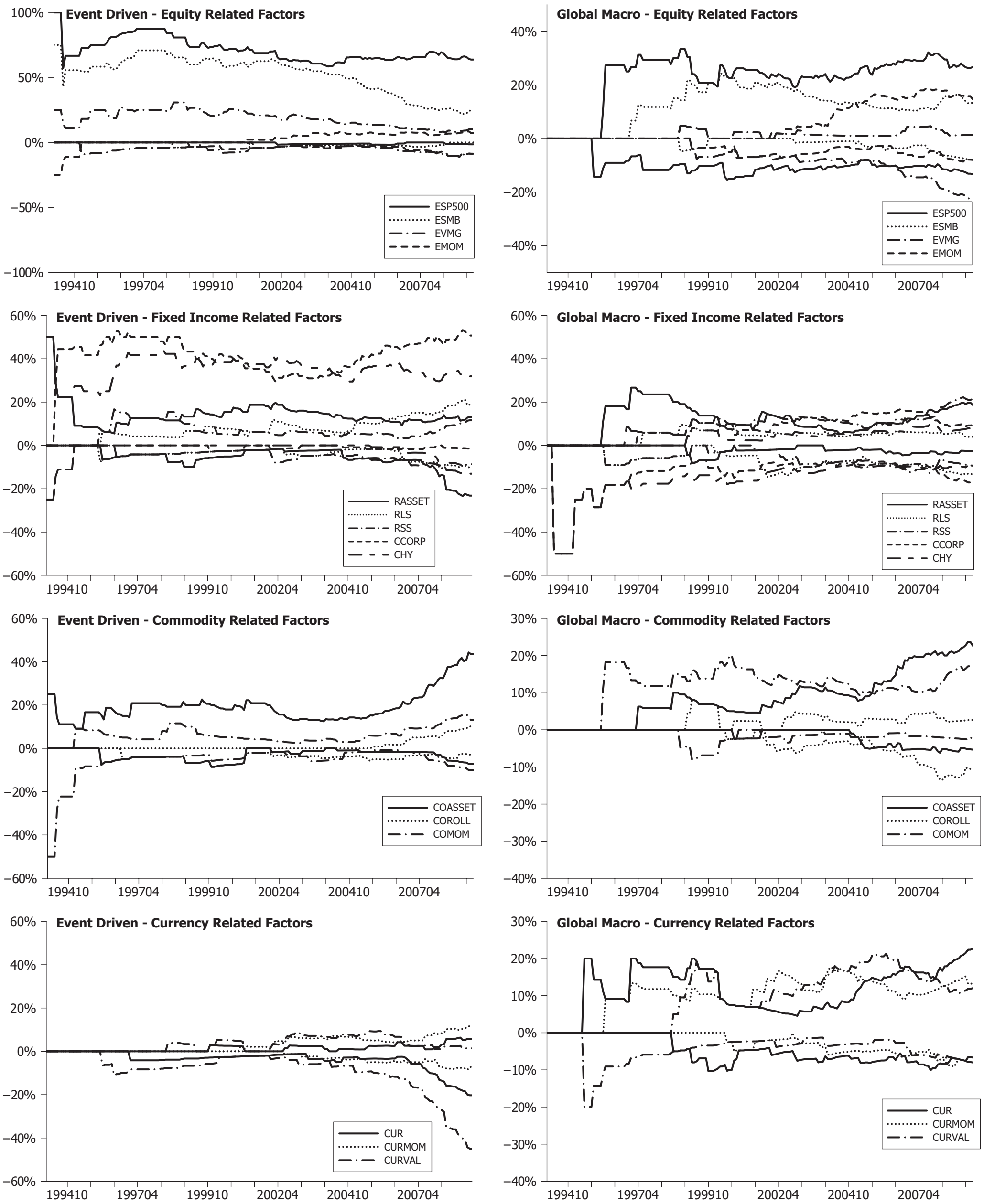


Figure 1.7: Factor Exposure Over Time (by Strategy Type - Multi Strategy and Single Strategy)

The following plots show the percentage of hedge funds with positive and negative exposure to the new risk factors proposed in this study for the Multi Strategy and Single Strategy strategy types over time. The four plots per hedge fund strategy type show the exposure to the equity-, fixed income-, commodity- and currency-related factors separately. The lines above the zero-mark display the percentage of funds with positive exposure and the lines below the zero-mark display the percentage of funds with negative exposure. The percentage exposure to the risk factors are derived from the structural change regression approach described in subsection 1.3.2.

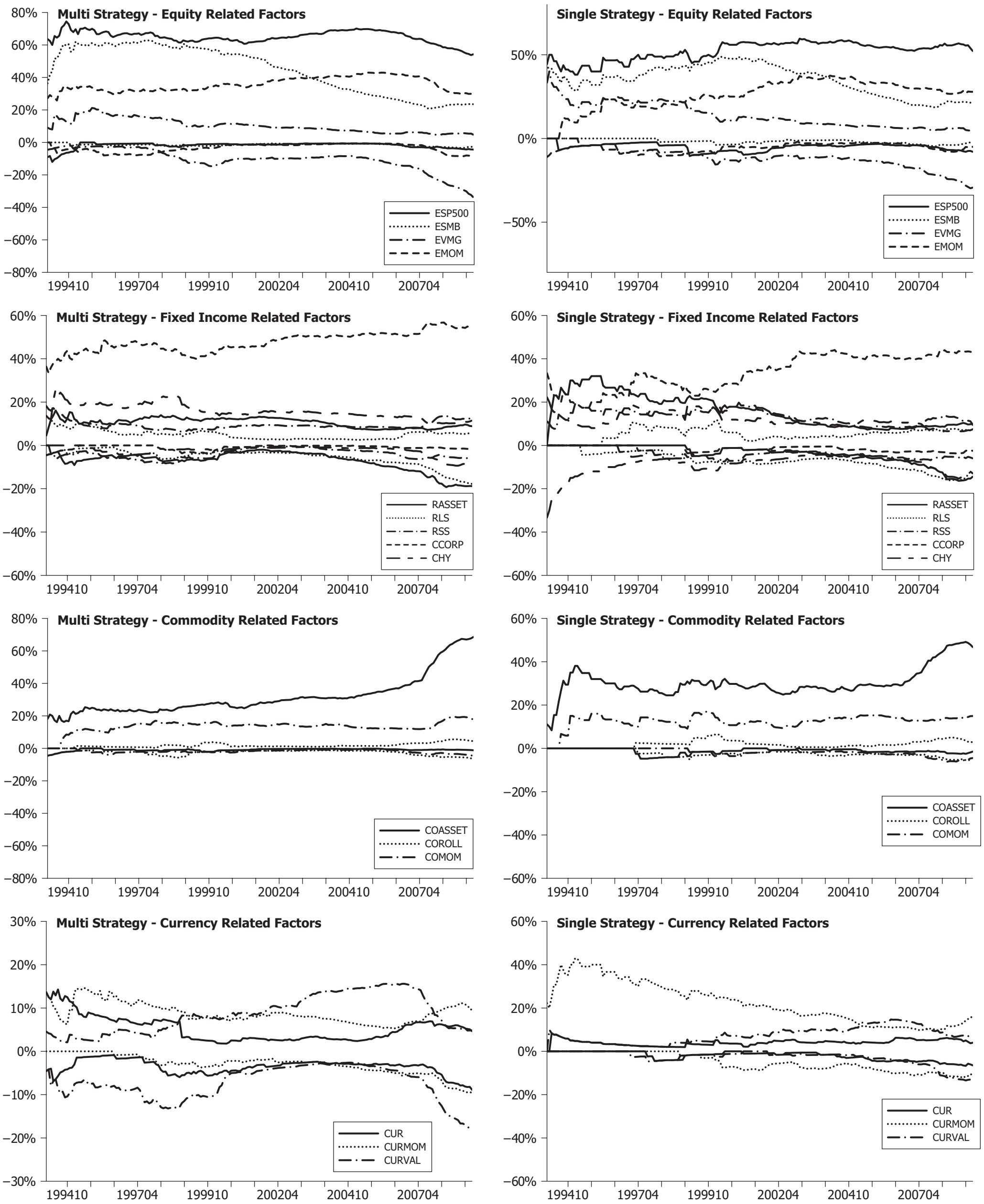
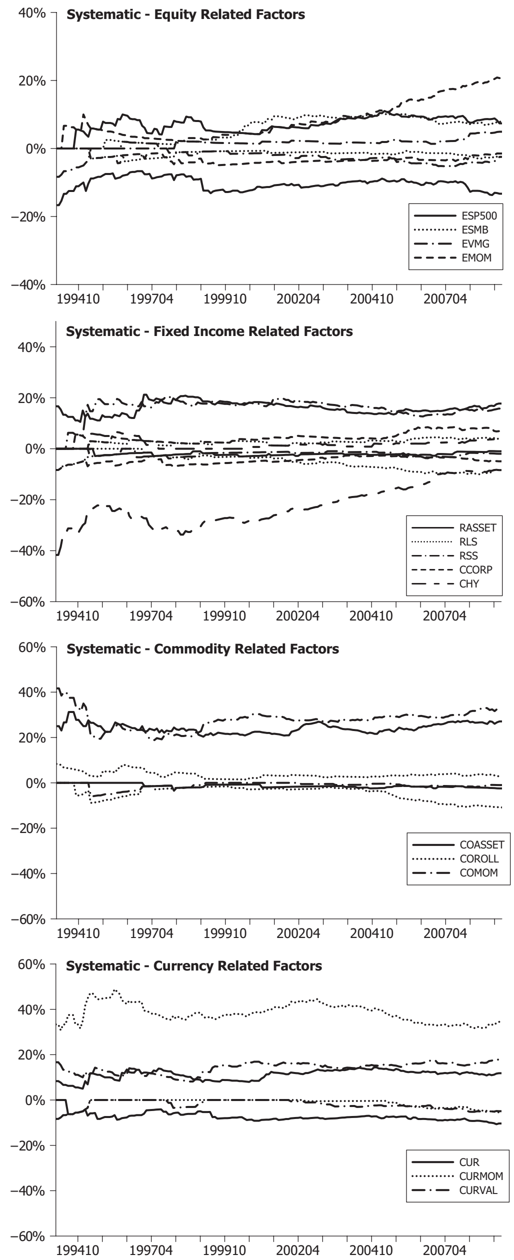


Figure 1.8: Factor Exposure Over Time (by Strategy Type - Systematic)
The following plots show the percentage of hedge funds with positive and negative exposure to the new risk factors proposed in this study for the Systematic strategy type over time. The four plots show the exposure to the equity-, fixed income-, commodity- and currency-related factors separately. The lines above the zero-mark display the percentage of funds with positive exposure and the lines below the zero-mark display the percentage of funds with negative exposure. The percentage exposure to the risk factors are derived from the structural change regression approach described in subsection 1.3.2.



The most important findings from the analysis of the factor exposures over time can be summarized as follows:

Equity Related Factors

The most interesting finding is related to the ESMB factor which represents the small cap risk premium in equities. For the HF, FoF, Event Driven, Multi Strategy and Single Strategy there is clear evidence that the exposure of hedge funds to this risk factor has continuously declined during the observed period. For example in the HF category, at the start of the data series almost 60% of the hedge funds had exposure to the ESMB factor. At the end of the time period, this value has dropped to 23%. For the EVMG risk factor, representing the value minus growth premium, the results show that in the FoF and Multi Strategy categories, an increasing percentage of funds have increased the negative exposure to this factor during the sub-prime crisis of 2008 and 2009. A more pronounced tilt towards a negative exposure to the EVMG factor is also observed after the LTCM crisis in 1998 until the beginning of the year 2000. With regards to the EMOM risk factor, representing the equity momentum risk premium, there is evidence that the risk factor over time has continuously become more important for the CTA, CPO, Systematic and Global Macro categories.

Fixed Income Related Factors

With regards to the fixed income related factors, the results show that the percentage of funds with exposure to the CCORP risk factor, representing credit risk exposure, has increased over time for the HF, FoF, Multi Strategy and Single Strategy categories. For the CTA, CPO and Systematic categories, the CHY factor, representing the risk premium of high yield exposure, seemed to have played an important role in the past. In all three categories more than 20% of the funds had negative exposure to the risk factor however, the percentage of funds with negative exposure clearly declined after 1998.

Commodity Related Factors

The positive exposure of the HF, FoF, Event Driven, Global Macro, Multi Strategy and Single Strategy categories to the COASSET risk factor, representing exposure to a broad commodity index risk premium, has increased in a remarkable way over the observed period. Thereby, the largest increase can be observed towards the end of the time series starting in 2007 which coincides with the start of the final rally in the commodity price boom which ended in June of 2008[37]. Thus, the analysis provides a clear indication that hedge funds in various categories have shifted their exposure to the commodity market during the commodity price boom.

Currency Related Factors

The most notable observation with regards to the currency factors is that the percentage of funds with exposure to the CUR factor, representing the carry-trade risk premium, and the percentage of funds with exposure to the CURVAL factor, representing the currency value premium, decreased during times of market turmoil in the FoF and Multi Strategy category. This pattern is seen in 1998 as well as during the sub-prime crisis of 2008 to 2009. The HF category shows similar patterns, but less pronounced. The Global Macro category shows a decrease of the funds with positive exposure and an increase of funds with negative exposure to the CUR factor during the time period of 1998 to 2000. During the sub-prime crisis however, the percentage of funds with positive exposure to the CUR factor actually increased. Overall, these results can be interpreted as reflecting a flight to quality in terms of currency exposure during market turmoil.

The results of the analysis of the factor exposure are of additional interest

[37] The S&P GSCI Total Return index increased by 88% from the beginning of January 2007 until the end of June 2008.

as they challenge the findings of Sandvik et al. (2011). In their study, they split a data set of hedge fund style indices for the time period of January 1994 to February 2009 into bull and bear market periods on an ad hoc basis. Their bull market sample consists of the time periods from January 1994 to August 2000 and October 2002 to October 2007. Their bear market sample consists of the time periods from September 2000 to September 2002 and November 2007 to February 2009. They run a stepwise regression model on each sub-sample for every hedge fund strategy index in their sample. One of their main findings is, that the SMB factor is only relevant in bull markets as in their results, the factor is not included for any of the hedge fund indices in the bear market sample. The result of my study points into a different direction, namely that the relevance of the small cap premium risk factor has steadily declined over time, rather than that it is linked to the market environment[38]. This finding is coherent with the recent discussion in the literature pointing to the direction that the size effect has disappeared since it was first documented by Banz (1981)[39]. A second finding of Sandvik et al. (2011) is that hedge fund managers tend to get more exposure to the commodity market during falling equity markets. The findings in this study do not support their conclusion as the results indicate that the rise in percentage of funds with commodity exposure is overlapping with the commodity price boom which started in early 2002 and accelerated at the beginning of 2007. In this study, I find no indication that the percentage of hedge funds with exposure to commodities was higher during the crisis of 1998 (LTCM) or 2002 (dot-com bubble).

[38] This result is unchanged when the constrained data sample of subsection 1.4.2 is used which drops the observations from July 2008 to March 2009.

[39] A recent discussion of the issue is found for example in Hou and Van Dijk (2010).

1.5 Conclusion

This study introduced a new set of factors which are based on risk premia for explaining the performance of hedge funds. The chosen regression approach is based on a best subset regression which chooses factors to be included in the model by using the Bayesian information criterion in order to minimize the problem of over-fitting. The new factors prove to be superior at explaining the average variation as well as the level of returns for all considered hedge fund categories when compared to a benchmark model using the straddle factors introduced in Fung and Hsieh (2001). In addition, the new factors have the advantage, compared to the benchmark model factors, that they are based on static and dynamic strategies which have become easily accessible for institutional investors. Thus, the new factors are not only superior at explaining the performance of hedge funds, but they can potentially also improve the hedge fund allocation and risk management processes of multi asset class investors.

Due to the strong evidence in the existing literature that the exposure of hedge funds to risk factors change over time, this study applied the new risk factors also in a dynamic setting, where the exposure of individual hedge funds is allowed to change at discrete points in time. By using a multiple structural change regression approach, which allows for multiple switches in exposure, this study extended the work of Bollen and Whaley (2009) which only allowed for a single switch in exposure. It is shown, that the structural change regression approach is superior at explaining the variation of hedge fund returns on average. On the other side, the average estimated alpha is actually higher when compared to a static best-subset regression approach. In addition to the average measures over the entire data sample, it was also analyzed how the performance of hedge funds developed over time. It is shown that the percentage of funds with an insignificant alpha has increased over time in the majority of the hedge fund categories under consideration, whereby the sub-prime crisis towards the end of the data sample had a significant impact. Also, the best performance of hedge funds is identified during the period of late 1998 to the beginning of 2000. Both results are coherent with the results in the existing literature. A further interesting finding of this study shows, that single hedge funds tend to switch their exposure at points in time where big shifts in the financial markets take place. This result is in-line with the findings in the existing literature about structural breaks in hedge fund return series, whereby the focus is mainly on hedge fund indices. However, the result contradicts the findings of Bollen and Whaley (2009) which found that switches in the exposure of hedge funds to risk factors does not generally occur at the points in time when big shifts in the markets occur. Thus, this study provides counter-evidence to the findings of Bollen and Whaley (2009) by showing that by allowing for multiple structural changes, the switching points found at the single hedge fund level are indeed similar to those found in the literature focused at hedge fund index returns.

The multiple structural break point approach also gives insights into how the exposure of single hedge funds to the different risk factors changes over time. The most interesting findings with regards to the factor exposure evolvement are that the equity small cap premium factor has steadily declined in importance over the observed time period and that the relevance of the commodity asset class factor increased remarkably during the final phase of the commodity price boom in the years 2007 and 2008. These findings contradict the conclusions of Sandvik et al. (2011), who find that hedge fund managers generally have more exposure to commodity markets in equity bear markets and that the small cap premium factor is only relevant in equity bull markets. This study found that the first result is only true for the sub-prime crisis in 2008 and that the relevance of the small cap premium has decreased over time, independent of the market cycles.

The results of this study provide evidence that using risk premia based risk factors can be superior at explaining the performance of hedge funds and that using a multiple structural break regression approach can give valuable insights into how the performance and positioning of hedge funds change over time. These insights provide the grounds for challenging certain findings in the existing literature. In further research, the risk factor set used in this study could be supplemented by additional factors. Particularly, risk factors related to volatility premia could be promising. Also, the multiple structural break regression approach could be applied on a broader data set of merged hedge fund databases in order to generate results which cover a larger part of the hedge fund market. In addition, it may make sense to apply the risk premia factor approach in other fields outside the hedge fund performance attribution literature. Interesting topics may be found in the portfolio choice literature.

1.6 Appendix

Data

Short Horizon Term Premium Factor (Rates Short-term Spread RSS)

The risk factor representing the exposure to the short horizon term premium (RSS), is constructed by using monthly settlement price data on the 3-month Eurodollar Futures[40]. The idea is that there is a premium for holding a long position in a 3-month Eurodollar Future as a speculative long position is a bet that interest rates in the future will be lower than what is anticipated by the market. Holding a long position in the 3-month Eurodollar Futures is therefore equivalent to taking on the risk of unexpected interest rate rises at the short end of the yield curve[41].The focus will be on contracts which expire 5-12 quarters in the future. The short horizon term premium risk factor

[40] Datastream Code: IEDmmyy(PS) where mm and yy refer to the month respectively year of the expiration of the contract.

[41] The settlement price is defined as 100 minus the fixing of the 3-month USD LIBOR rate.

(RSS) is constructed in two steps. First, for each of the considered maturities, indices are constructed which represent the return of rolling the respective 3-month Eurodollar Futures every 3 month. For example, in January 2000 the index representing the contracts with a time to maturity of 5 quarters would have full notional exposure to the 3-month Eurodollar Futures contract expiring in March 2001. In March 2000 this position would be rolled into the 3-month Eurodollar Futures expiring in June 2001 etc. Similarly, in January 2000, the index representing the contracts with a time to maturity of 6 quarters, would have full notional exposure to the 3-month Eurodollar Futures expiring in June 2001 and would roll the exposure into the September 2001 contract in March 2000. This procedure is applied to all 8 indices representing the respective Eurodollar futures with times to maturity of 5 to 12 quarters. In a second step, the RSS risk factor is then created by having equal weight exposure to all 8 indices.

Commodity Momentum Factor (COMOM)

In order to create the commodity momentum factor (COMOM), in a first step, investable indices for each commodity are calculated which I will refer to as single commodity indices. The monthly return for each single commodity index is determined by the return of the futures contract which settles in two months time. At then end of each month the position is rolled into the next contract. For example, the return of the WTI crude oil commodity index for the month of March is equal to the return on a May WTI futures contract for that month[42]. The return for the month of April is subsequently equal to the return on a June WTI futures contract during the month of April. This procedure ensures that gains and losses from rolling commodity future contracts are fully accounted for in the single commodity index. The single commodity indices are then used to construct a long-short momentum

[42] The last trading day of the May WTI futures contract is in April. Detailed information on the expiration schedule for WTI crude oil futures contracts can be found for example on the homepage of the CME group: https://www.cmegroup.com

portfolio. I follow the approach of Miffre and Rallis (2007) by ranking the single commodity indices according to the performance in the ranking period which I choose to be one month. Thus, at the end of every month, the single commodity indices are ranked according to their 1 month performance from the best performing commodity to the worst performing commodity. Then, the long leg of the momentum portfolio is determined by giving equal (positive) weight to the single commodity indices in the top quintile and the short leg is determined by giving equal (negative) weight to the single commodity indices in the bottom quintile[43]. The holding period of the long-short portfolio is chosen to be 3 months. This leads to a minor issue as the optimal portfolios are overlapping (e.g after one month time, a new optimal portfolio can be constructed while the optimal portfolio from the prior month still needs to be held for another two months). As in Miffre and Rallis (2007), I circumvent the issue by equally weighting every month the top and bottom 3 quintile portfolios that were formed during the last three month. Thus, for example at the end of April, I equally weight the long and short portfolios derived at the end of the current month, the end of March and the end of February. This leads to an aggregated portfolio which consists of three long-short sub portfolios which are monthly rebalanced. This aggregated portfolio constitutes the commodity momentum risk factor[44].

Commodity Roll Yield Factor (COROLL)

Similar to the study of Fuertes, Miffre and Rallis (2010) when constructing the commodity roll yield factor (COROLL), I also determine first how much the available commodities are in backwardation respectively contango. Thereby, I examine the prices of the futures contract which settle in two

[43]The weights assigned to the long leg sum to +1, and the weights assigned to the short leg sum up to -1. Thus the portfolio has a net exposure of zero.

[44]The risk return characteristics of the derived commodity momentum risk factor are close to the ones derived in Fuertes, Miffre and Rallis (2010) when comparing the results to the corresponding long-short momentum portfolio for the matching time period (January 1979 - January 2007).

months time which were also used for constructing the commodity momentum factor. I refer to this contract as the first contract and denote its price as P_1. In addition, I also use the price of the next contract which expires directly after the first contract and denote its price as P_2. As in Fuertes, Miffre and Rallis (2010) I use the following formula at each point in time t, to determine the implied roll return $R_{t,i}$ of each commodity i. Thereby, $N_{t,i,2}$ and $N_{t,i,1}$ are the number of days until maturity on date t for the second and the first contract respectively.

$$R_{t,i} = (ln(P_{t,i,1}) - ln(P_{t,i,2})) \times 365/(N_{t,i,2} - N_{t,i,1}) \tag{1.4}$$

Knowing the implied roll yield $R_{t,i}$ of each commodity at point t, I determine the top and bottom quintile of the most backwardated and most contangoed commodities. Instead of simply taking long positions in the most backwardated commodities and short positions in the most contangoed commodities, I create long-short portfolios in each commodity. For each of the commodities in the most backwardated quintile, the strategy goes long the second contract and shorts an amount of equal notional value of the first contract. Similarly, for each commodity in the most contangoed quintile, the strategy goes long the first contract and short the second contract. The idea behind this approach is to isolate the pure roll yield by hedging out changes in the level of the commodity term structure. The return $r_{t+1,COROLL}$ of the commodity roll yield factor for the time period from t to $t+1$ is then determined by equally weighting the long-short positions taken in the most backwardated and the most contangoed commodities. The formula for calculating $r_{t+1,COROLL}$ is given below:

$$r_{t+1,COROLL} = \sum_{i}^{Q1_t} w_{t,q1} \left(\left(\frac{P_{t+1,i,2}}{P_{t,i,2}} - 1 \right) - \left(\frac{P_{t+1,i,1}}{P_{t,i,1}} - 1 \right) \right) \tag{1.5}$$

$$+\sum_{j}^{Q5_t} w_{t,q5}\left(\left(\frac{P_{t+1,j,1}}{P_{t,j,1}}-1\right)-\left(\frac{P_{t+1,j,2}}{P_{t,j,2}}-1\right)\right)$$

Thereby, $Q1_t$ and $Q5_t$ are the number of commodities in the top and bottom quintile respectively at time t and $w_{t,q1} = 1/Q1_t$ and $w_{t,q5} = 1/Q5_t$. The resulting commodity roll yield risk factor COROLL has similar risk-return properties as the strategy introduced by Fuertes, Miffre and Rallis (2010), while exhibiting a lower absolute volatility and a lower correlation with the commodity momentum factor when compared to their strategy[45].

Currency Carry Factor (CUR)

The currency carry trade strategy is implemented by first determining the implied 1-month interest rate differential for each currency i at time t by the following formula, where $F_{t,i}^{mid}$ and $S_{t,i}^{mid}$ refer to the mid 1-month forward rate and the mid spot rate at time t for currency i.[46]:

$$sigCarry_{t,i} = \frac{F_{t,i}^{mid}}{S_{t,i}^{mid}} - 1 \tag{1.6}$$

The used forward and spot foreign exchange rates are quoted in units of USD per one unit of foreign currency[47]. Therefore, if the interest rate differential as determined by the above equation is <0, this implies that the foreign currency carries a higher interest rate than the USD and should therefore depreciate. In order to profit from the higher interest rate in the foreign currency and take a bet that it will depreciate by less than is implied by the interest rate differential at time t, an investor can enter into a long 1-month forward contract at the rate $F_{t,i}^{Ask}$. After 1-month time, at $t+1$, the investor has to deliver dollars in the amount of $F_{t,i}^{Ask}$ and receives one unit of foreign

[45] Replacing the COROLL risk factor by a factor which mimics the term structure strategy introduced by Fuertes, Miffre and Rallis (2010) does not change the results of this study.

[46] The mid rates are calculated by taking the arithmetic average of the bid and ask prices.

[47] Thus, they represent the price a USD investor has to pay in order to buy one unit of foreign currency.

currency. The investor makes a profit on this trade whenever the spot rate $S^{Bid}_{t+1,i}$, at which he can exchange the unit of foreign currency back into USD, is higher than the forward rate $F^{Ask}_{t,i}$. The return $r_{t+1,i,long}$ in terms of the USD notional value invested in the forward contract at time t is therefore determined by:

$$r_{t+1,i,long} = \frac{S^{Bid}_{t+1,i}}{F^{Ask}_{t,i}} - 1 \tag{1.7}$$

Similarly, if the interest rate differential is >0 the investor would enter into a short 1-month forward contract and the return in terms of the USD notional value invested in the forward contract would equal to:

$$r_{t+1,i,short} = -\frac{S^{Ask}_{t+1,i}}{F^{Bid}_{t,i}} - 1 \tag{1.8}$$

The currency carry risk factor CUR is constructed by first determining the interest rate differential for all 15 currencies versus the USD at the end of each month. Then, the long leg of the portfolio is defined by an equally weighted portfolio of long forward contracts for those currencies for which the interest rate differential is negative. The short leg is created by an equally weighted portfolio of short forward contracts for the currencies for which the interest rate differential is positive. The returns in each currency bet for one period are calculated by applying the formulas above on each long respectively short position. The formula for determining the one period return for the aggregated trade which represents the CUR risk factor is shown below:

$$r_{t+1,CUR} = \sum_{i}^{N_{t,long}} w_{t,long} r_{t+1,i,long} + \sum_{i}^{N_{t,short}} w_{t,short} r_{t+1,i,short} \tag{1.9}$$

The parameters $N_{t,long}$and $N_{t,short}$ are the number of currencies for which a long position respectively a short position in the forward contract is opened at time t. The weights $w_{t,long}$ and $w_{t,short}$ are defined as $1/N_{t,long}$ and $1/N_{t,short}$ respectively.

Currency Momentum Factor (CURMOM)

The currency momentum factor CURMOM and the currency value factor CURVAL are constructed in the same way as the currency carry factor CUR. The only difference is the definition of the signal which determines at the end of each month whether a long or short position is taken in the forward contracts. For the momentum factor, the signal generation is simple. At the end of each month a momentum signal is generated which is equal to the spot rate return over the past three months:

$$sigMom_{t,i} = \frac{S_{t,i}^{mid}}{S_{t-3,i}^{mid}} - 1 \tag{1.10}$$

The strategy assumption is, that currencies which have appreciated during the past 3-month, will continue to do so next month, while those which have depreciated will also continue on this path for the next month. Thus, as for the currency carry factor, the long leg of the portfolio is constructed by an equally weighted portfolio of long forward contracts for the currencies for which the momentum signal is positive. Similarly, a portfolio of short positions is created for the currencies for which the momentum signal is negative. The return $r_{t+1,CURMOM}$ for the currency momentum risk factor is then calculated as in equation 1.9, with the only difference that the constituents in the long and the short leg are determined by the momentum signal.

Currency Value Factor (CURVAL)

The last currency factor is the value factor. The idea behind this factor, as put forward for example in Kroencke, Schindler and Schrimpf (2011), is that currencies trade above or below the exchange rate which is justified by the fundamental value of the currency as for example measured by the purchasing power parity (PPP). An investor exploiting a value strategy in currencies would therefore go long currencies which are undervalued and go short currencies which are overvalued from a fundamental point of view. In this study, I use the effective exchange rate (EER) broad indices published by the bank for international settlements (BIS) to determine the relative value of the currencies in my data set[48]. Similar to the previous two currency strategies, a signal is generated which indicates whether a currency is over- or undervalued against the USD and therefore determines whether a long or short position should be entered in the respective forward contracts[49]. The signal is generated by taking the difference between the two month lagged effective exchange rate $EER_{t-2,i}$ of the foreign currency i and the lagged effective exchange rate of the USD $EER_{t-2,USD}$[50].

$$sigVal_{t,i} = EER_{t-2,i} - EER_{t-2,USD} \tag{1.11}$$

Again, as for the previous currency risk factors, at the end of each month the long leg of the portfolio is being constructed by an equally weighted portfolio of long forward contracts for the currencies which are undervalued. Similarly, an equally weighted portfolio of short forward contracts is defined for the currencies which are overvalued according to the signal defined in equation 1.11. The return $r_{t+1,CURVAL}$ is calculated as in equation 1.9 where the constituents of the long and short leg are determined by the value signal.

[48] The data can be downloaded under the following link: http://www.bis.org/statistics/eer/index.htm.

[49] All currencies listed in table 1.15 are considered for the CURVAL factor except the Kuwait-Dinar (KWD), for which no effective exchange rate data is available.

[50] The lag is introduced to ensure that only data is being used which could have been available at the time the portfolio is being rebalanced.

Table 1.13: CISDM Database Fund Strategies

List of strategy categories of the CISDM database. The second column refers to the number of funds which fulfill the inclusion criteria described in subsection 1.2.1.

Strategy Type	Number of Funds
Capital Structure Arbitrage	14
Conservative	14
Convertible Arbitrage	94
Discretionary	88
Distressed Securities	93
Emerging Markets	229
Equity Long Only	77
Equity Long/Short	953
Equity Market Neutral	133
Event Driven Multi Strategy	146
Fixed Income	85
Fixed Income - MBS	53
Fixed Income Arbitrage	71
Global Macro	137
Invest Funds in Parent Company	15
Market Neutral	61
Market Timing	1
Merger Arbitrage	33
Multi Strategy	1272
Opportunistic	25
Option Arbitrage	22
Other Relative Value	11
Regulation D	1
Relative Value Multi Strategy	53
Sector	190
Short Bias	14
Single Strategy	385
Systematic	350

Table 1.14: Commodities Data Summary

List of the commodity futures used for the calculation of the COMOM and COROLL risk factors. The *Availability of Data* column shows the month from which onwards the respective commodity was included in the calculation of the risk factors.

Commodity Name	Datastream Ticker	Availability of Data
CrudeOil	NCL	Jan 1994
BrentCrudeOil	LLC	Sep 2003
UnleadedGas	NRB	Oct 2005
HeatingOil	NHO	Jan 1994
GasOil	LLE	Sep 2003
NaturalGas	NNG	Jan 1994
Aluminium	LAH	Jul 1993
Copper	NHG	Jan 1994
Lead	LED	May 2002
Nickel	LNI	May 2002
Zinc	LZZ	May 2002
Gold	NGC	Jan 1994
Silver	NSL	Jan 1994
Wheat	CW.	Jan 1994
Corn	CC.	Jan 1994
Soybeans	CS.	Jan 1994
Cotton	NCT	Jan 1994
Sugar11	NSB	Jan 1994
Sugar14	NSE	Jan 1994
Coffee	NKC	Jan 1994
Cocoa	NCC	Jan 1994
LiveCattle	CLC	Jan 1994
FeederCattle	CFC	Jan 1994
LeanHogs	CLH	Jan 1994
OrangeJuice	NJO	Jan 1994
SoybeanMeal	CSM	Jan 1994
SoybeanOil	CBO	Jan 1994
WheatKansasCity	KKW	Jan 1994
WheatHardWinter	MIJ	Dec 2004
FrozenPorkBellies	CPB	Jan 1994
Palladium	NPA	Jan 1994
Platinum	NPL	Jan 1994
Tin	LTI	May 2002
Milk	CFM	Mar 1996
Lumber	CLB	Jan 1994

Table 1.15: Currencies Data Summary

List of the currency spot and 1-month forward prices used for the calculation of the CUR, CURMOM and CURVAL risk factors. The bid and offer rates are available by adding the suffix (EB) and (EO) respectively to the Datastream tickers. For the calculation of the risk factors, all exchange rates are converted to represent the price of one unit of foreign currency in USD.

Currency Pair	Datastream Ticker Spot Price	Datastream Ticker 1-M Forward Price
USDAUD	TDAUDSP	TDAUD1F
CADUSD	TDCADSP	TDCAD1F
CHFUSD	TDCHFSP	TDCHF1F
DKKUSD	TDDKKSP	TDDKK1F
USDEUR	TDEURSP	TDEUR1F
USDGBP	TDGBPSP	TDGBP1F
HKDUSD	TDHKDSP	TDHKD1F
JPYUSD	TDJPYSP	TDJPY1F
KWDUSD	TDKWDSP	TDKWD1F
NOKUSD	TDNOKSP	TDNOK1F
USDNZD	TDNZDSP	TDNZD1F
SARUSD	TDSARSP	TDSAR1F
SEKUSD	TDSEKSP	TDSEK1F
SGDUSD	TDSGDSP	TDSGD1F
ZARUSD	TDZARSP	TDZAR1F

Table 1.16: Summary Statistics of Static Regression (January 1994 - June 2008)

The table lists the adjusted R-squared, the average alpha, the percentage of funds with an insignificant alpha at the 95% confidence level (one-sided interval) as well as the number of factors which are included in the model for the new factor model and the benchmark factor model, each defined in subsection 1.2.2. The regressions are conducted by a best subset approach where the factors included in the model are selected using the Bayesian Information Criterion (BIC). The confidence level is calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors, whereby the automated lag selection described in Newey and West (1994) is applied. Compared to the results in table 1.4, these results were found by using a constrained data sample covering the period from January 1994 to June 2008 only.

	New Factor Model				Benchmark Factor Model			
	adj R2	alpha	insignificant alpha	factors	adj R2	alpha	insignificant alpha	factors
HF	0.35	0.0055	0.52	2.85	0.26	0.0070	0.43	1.82
CTA	0.22	0.0058	0.69	2.24	0.18	0.0098	0.50	1.59
FoF	0.44	0.0024	0.60	3.78	0.36	0.0043	0.33	2.89
CPO	0.24	0.0036	0.76	2.55	0.20	0.0076	0.53	1.99
Event Driven	0.38	0.0055	0.43	2.91	0.33	0.0070	0.36	2.25
Global Macro	0.27	0.0055	0.61	2.37	0.18	0.0068	0.54	1.56
Multi Strategy	0.39	0.0035	0.56	3.42	0.31	0.0057	0.29	2.59
Single Strategy	0.39	0.0037	0.59	3.30	0.29	0.0056	0.42	2.14
Systematic	0.23	0.0043	0.73	2.34	0.19	0.0090	0.47	1.69
Average	0.32	0.0044	0.61	2.86	0.25	0.0070	0.43	2.06

Table 1.17: Summary Statistics of Static Regression (Enhanced Benchmark Model)

The table lists the adjusted R-squared, the average alpha, the percentage of funds with an insignificant alpha at the 95% confidence level (one-sided interval) as well as the number of factors which are included in the model for the new factor model and an enhanced benchmark factor model. The enhanced benchmark model includes the following three additional risk factors which have been used in the existing literature: EVMG, CHY and COASSET. The regressions are conducted by a best subset approach where the factors included in the model are selected using the Bayesian Information Criterion (BIC). The confidence level is calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied. The used data sample covers the period from January 1994 to March 2009.

	New Factor Model				Benchmark Factor Model			
	adj R2	alpha	insignificant alpha	factors	adj R2	alpha	insignificant alpha	factors
HF	0.40	0.0049	0.57	3.17	0.36	0.0062	0.48	2.59
CTA	0.24	0.0053	0.71	2.40	0.22	0.0081	0.57	2.00
FoF	0.52	0.0014	0.74	4.03	0.50	0.0031	0.51	3.74
CPO	0.25	0.0032	0.78	2.72	0.24	0.0064	0.61	2.57
Event Driven	0.44	0.0053	0.45	3.39	0.42	0.0064	0.38	2.99
Global Macro	0.28	0.0053	0.58	2.51	0.25	0.0067	0.50	2.20
Multi Strategy	0.48	0.0027	0.68	3.82	0.45	0.0044	0.44	3.51
Single Strategy	0.46	0.0027	0.67	3.59	0.42	0.0043	0.51	3.02
Systematic	0.26	0.0039	0.75	2.50	0.23	0.0075	0.54	2.11
Average	0.37	0.0039	0.66	3.13	0.34	0.0059	0.50	2.75

Empirical Results

Table 1.18: Significance of Alphas Using Bootstrap Method
The table lists the percentage of funds with an insignificant alpha at the 95% confidence level whereby the bootstrap percentile method as it is described for example in Efron (1981) was used to construct the confidence intervals. The regression approach corresponds to the one described in table 1.4 of subsection 1.4.1. In order to determine the confidence intervals, for each hedge fund category and factor model setup 500 bootstrap replications were used.

	New Factor Model insignificant alpha	Benchmark Factor Model insignificant alpha
HF	0.68	0.55
CTA	0.82	0.63
FoF	0.80	0.55
CPO	0.85	0.65
Event Driven	0.61	0.40
Global Macro	0.74	0.67
Multi Strategy	0.75	0.49
Single Strategy	0.76	0.53
Systematic	0.87	0.64
Average	0.76	0.57

Structural Change Regression Analysis

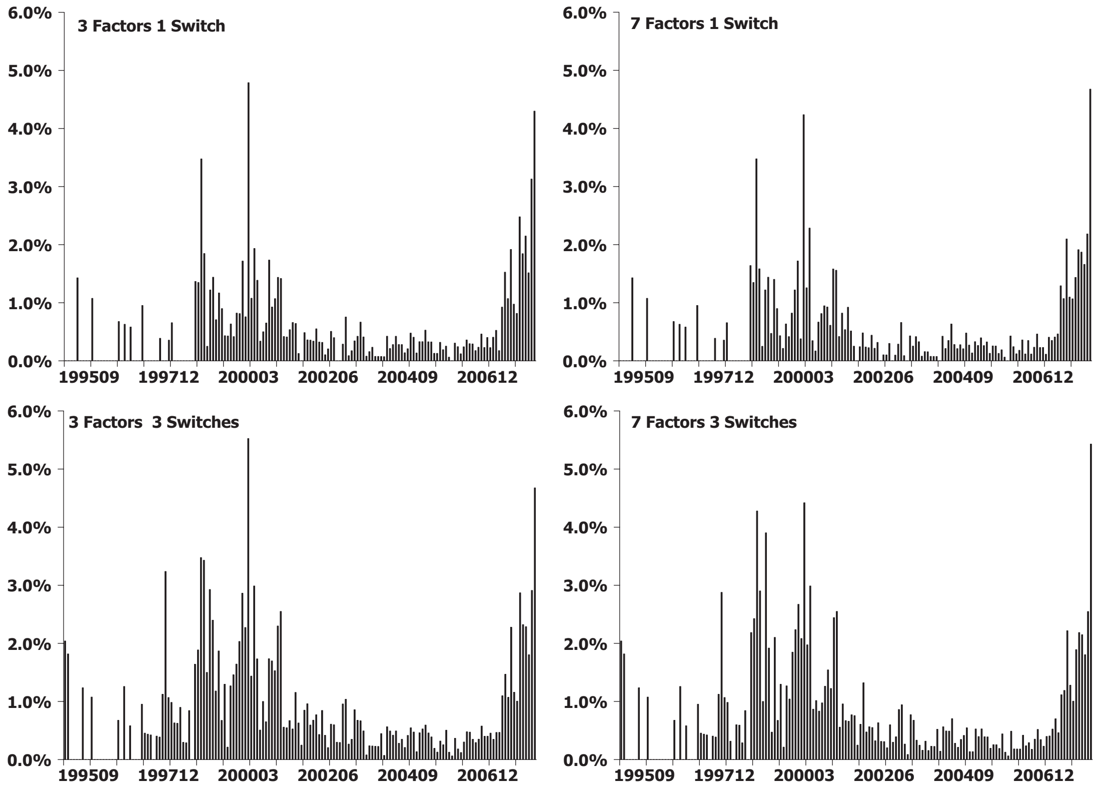
The approach and methodology used in this study differ in various parts from the one taken by Bollen and Whaley (2009). Therefore, the differences in the results concerning the switching frequency of funds over time of the two studies may stem from different sources. There are five main differences in which the approach in this study differs from the one in Bollen and Whaley (2009). The first one is the used data-set, where both studies use the CISDM database, but use different filtering criteria for the funds which are included in the analysis. Compared to Bollen and Whaley (2009), the data-set in this study covers the longer time period from January 1994 until March 2009, while their data-set covers the period from January 1994 until December 2005. Compared to their paper, this study only includes hedge funds for which assets under management are being reported and exceed 5 million USD. Further, in their analysis on the switching frequency, they only include funds which were still live in December 2005, while this study includes live and dead funds. The second difference lies in the estimation methodology for finding switching points. This study uses a multiple structural break point regression which relies on the Bayesian information criterion for determining whether a breakpoint is being considered to be feasible or not. Bollen and Whaley (2009) on the other side use an Avg-F Test approach as described in Andrews, Lee and Ploberger (1996). The third difference is related to the estimation procedure where in this study, I allow for a maximum of three switching points, while the approach of Bollen and Whaley (2009) only allows for one switching point. The fourth difference relates to the factor set, where in this study a new factor set is being used compared to the Bollen and Whaley (2009) study. The fifth and last difference is concerning the estimation of the best subset regression, where in this study a maximum of seven factors are allowed while in the study of Bollen and Whaley (2009) only three factors are allowed.

The following analysis shows how variations in the setup of the approach, in the five areas explained above, alter the results concerning the switching frequency of hedge funds over time. Thereby, I will focus on the HF category only, since it is the largest category. The following results are similar when examining the FoF category but for the reason of brevity only the results for the HF category are reported[51]. In a first step, it is being analyzed whether the number of switches and the number of factors allowed in the best subset regression model have an impact on the switching distribution of hedge funds over time. As is seen in figure 1.9 the results do not differ in a substantial way when varying between allowing a maximum of three or seven factors in the model. On the other side, the number of switches does have an impact in the sense that more switches are identified and therefore the switching activity around the identified market events is more pronounced when allowing for three switches instead of one switch. This can be seen for example when comparing the percentage of switching funds during the LTCM crisis in late

[51]For the CTA and CPO categories the analysis of this section is redundant as I found very similar results as Bollen and Whaley (2009) did in their work.

1998 and the dot-com bubble crisis in early 2002. However, neither variations in the switching frequency nor the number of factors allowed in the model does materially change the assessment that switches in the factor exposures is more likely to occur during times of great shifts in the financial markets.

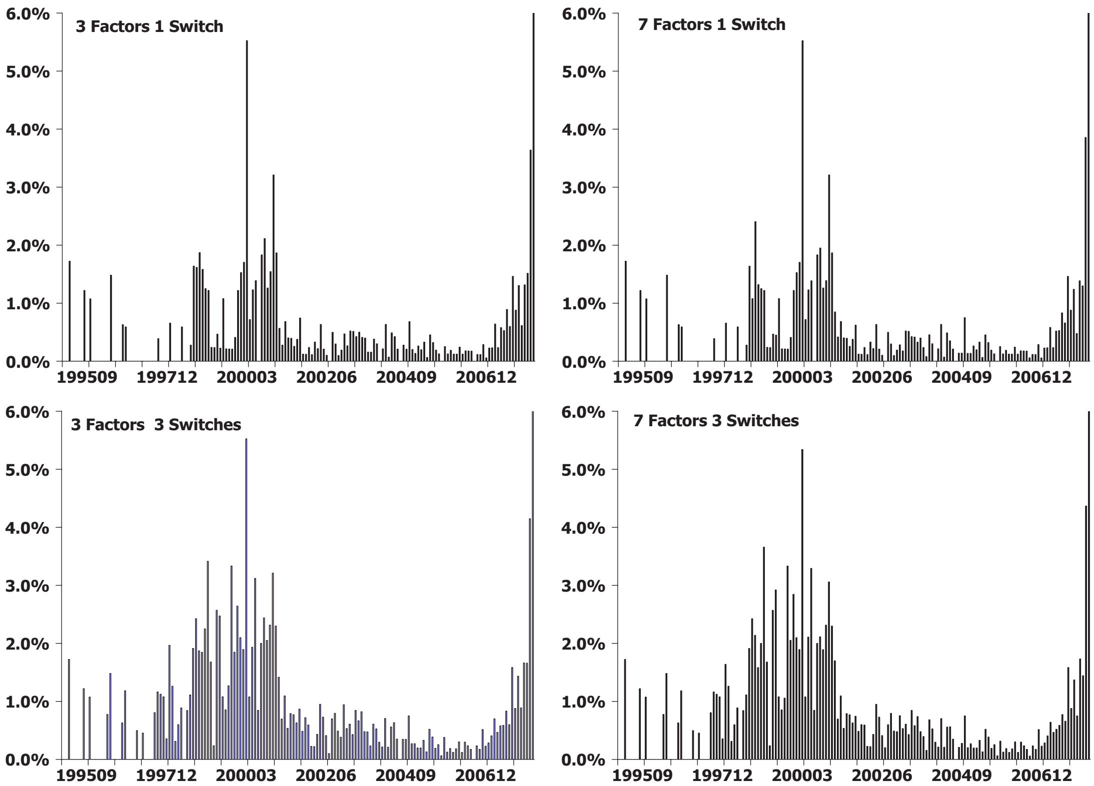
Figure 1.9: Switching Frequency - Variations New Factor Model Setup
The bar plots show for the HF hedge fund category the percentage of hedge funds at each point in the data sample for which a switch in factor exposure is identified using the structural change regression approach described in subsection 1.3.2. Thereby, the maximum number of factors allowed in the best subset regression is set to either three or seven and the number of maximum switches per fund are set to either one or three. The percentage measure is defined by the number of switching funds divided by the number of funds in the sample for which switching the exposure was feasible at the given month.



In a next step, it is being analyzed whether the used factor set has an impact on the distribution of the switching frequency over time. Thereby, the benchmark factor model which is similar to the one used by Bollen and Whaley (2009), is compared to the new proposed factor model. The graphs in figure 1.10 are using the same input parameters as those in figure 1.9 with the only difference that the factors of the benchmark model instead of the new factors are used. As can be seen in figure 1.10, the most prominent effect of using the benchmark factors is related to much fewer switches occurring around the date of the LTCM crisis in late 1998. This effect is more pronounced when only allowing for one switch in the model. Thus, there is strong evidence that the choice of the factors used in the factor model can have a material influence on the identification of switching points in the time series of hedge fund returns. At least in part the difference between the results of this study and the one of Bollen and Whaley (2009) can therefore be attributed to the use of different factor sets.

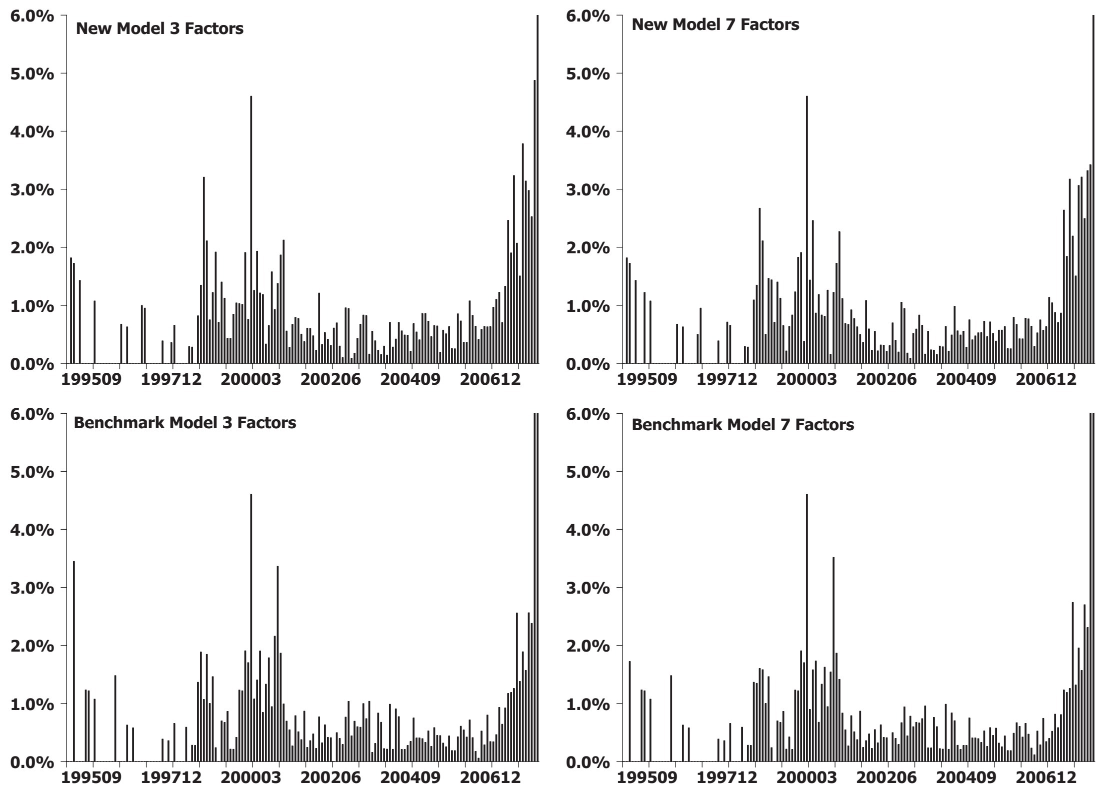
Figure 1.10: Switching Frequency - Variations Benchmark Factor Model Setup

The bar plots show for the HF hedge fund category the percentage of hedge funds at each point in the data sample for which a switch in factor exposure is identified using the structural change regression approach described in subsection 1.3.2, whereby instead of using the new introduced factors, the benchmark model factors are used. Thereby, the maximum number of factors allowed in the best subset regression is set to either three or seven and the number of maximum switches per fund are set to either one or three. The percentage measure is defined by the number of switching funds divided by the number of funds in the sample for which switching the exposure was feasible at the given month.



In a third step, the influence of the estimation procedure is being analyzed. The graphs in figure 1.11 are showing the switching frequency over time when an Avg-F Test is being used to identify switching points. Thereby, the identified switching point needs to be significant at the 90% level of confidence in order to be considered a feasible switching point. As can be seen, the results do not differ much from those in the previous figures where only one switch per fund is allowed. Thus, the choice of the methodology used to identify break points does not seem to materially affect the results when only allowing for one break point in the approach of Bai and Perron (2003). However, as seen in figures 1.9 and 1.10, allowing for multiple switches can have an impact on the distribution of the switching frequency over time.

Figure 1.11: Switching Frequency - Avg-F Test Approach
The bar plots show for the HF hedge fund category the percentage of hedge funds at each point in the data sample for which a switch in factor exposure is identified using the Avg-F-Test approach described in Andrews, Lee and Ploberger (1996). The analysis is conducted for the new factor model as well as for the benchmark factor model. Thereby, the maximum number of factors allowed in the best subset regression is set to either three or seven. The percentage measure is defined by the number of funds for which a switching point was identified and significant at the 90% level of confidence, divided by the number of funds in the sample for which switching the exposure was feasible at the given month.



The last variation in the setup which is being analyzed is related to the used data-set where different filtering criteria are used in this study compared to the one of Bollen and Whaley (2009). The graphs in figure 1.12 show the switching distribution when the Avg-F-Test approach applied in figure 1.11 is applied to a data-set were no constraints regarding the assets under management are imposed and which only includes funds which were live on March 2009. As is seen, the resulting switching distribution has relatively little mass around the date of the LTCM crisis, especially when the benchmark factor model is being used.

In summary, the analysis of the sensitivity of the switching distribution to the different parameter setups of the estimation approach showed, that the set of used factors does have a considerable impact on the distribution of the identified breakpoints. In addition, allowing for more than one break point is also relevant for the shape of the distribution. The new factor set as well as allowing for multiple switches, are therefore considered to be the main drivers of the diverging results of this study when compared to the study of Bollen and Whaley (2009). Further, as seen in the last part of the analysis, only including live funds is likely to be an additional source for varying results. On the other side, the estimation approach per se (Avg-F-Test approach vs. the structural regression approach) as well as the number of factors allowed in the model does not materially impact the results of the switching frequency of hedge funds.

To conclude this subsection, the regression analysis of section 1.4.2 is repeated whereby the original model is compared with the 3 factor benchmark model using the Avg-F-Test to identify switches in the factor exposure of funds[52]. As is shown in table 1.19, for the model using the benchmark factors with the Avg-F-Test methodology only the TED spread is significant at the 95% confidence level for explaining the switching frequency of funds in the HF category. Also, the R-squared measure is lower compared to the model

[52] This corresponds to the model with the switching frequency distribution shown in the lower left graph in figure 1.6.

introduced in this study. The difference between the two models is even more apparent when the same analysis is conducted on a constrained data-set, which uses only the data-points from January 1994 until December 2005 which corresponds to the time span analyzed by Bollen and Whaley (2009). As shown in table 1.20, during the shorter time period neither the VIX index nor the TED spread is found to be significant at the 95% confidence level for explaining the switching frequency resulting from the benchmark factor model using the Avg-F-Test. At the same time both factors are still highly significant for the new factor model when the shortened data sample is being used. The difference in the two models is also reflected in the adjusted R-squared measure which is considerably lower for the benchmark model when analyzing the shorter data sample.

Figure 1.12: Switching Frequency - Avg-F-Test Approach (only live funds)
The bar plots show for the HF hedge fund category the percentage of hedge funds at each point in the data sample for which a switch in factor exposure is identified using the Avg-F-Test approach described in Andrews, Lee and Ploberger (1996). Compared to the analysis shown in figure 1.11, the filtering criteria of the data-set have been changed such that only funds where included in the data-set which were live in March 2009 and no constraints regarding the assets under management were imposed. The adjusted data-set consists of 1804 hedge funds. The analysis is conducted for the new factor model as well as for the benchmark factor model. Thereby, the maximum number of factors allowed in the best subset regression is set to either three or seven. The percentage measure is defined by the number of funds for which a switching point was identified and significant at the 90% level of confidence, divided by the number of funds in the sample for which switching the exposure was feasible at the given month.

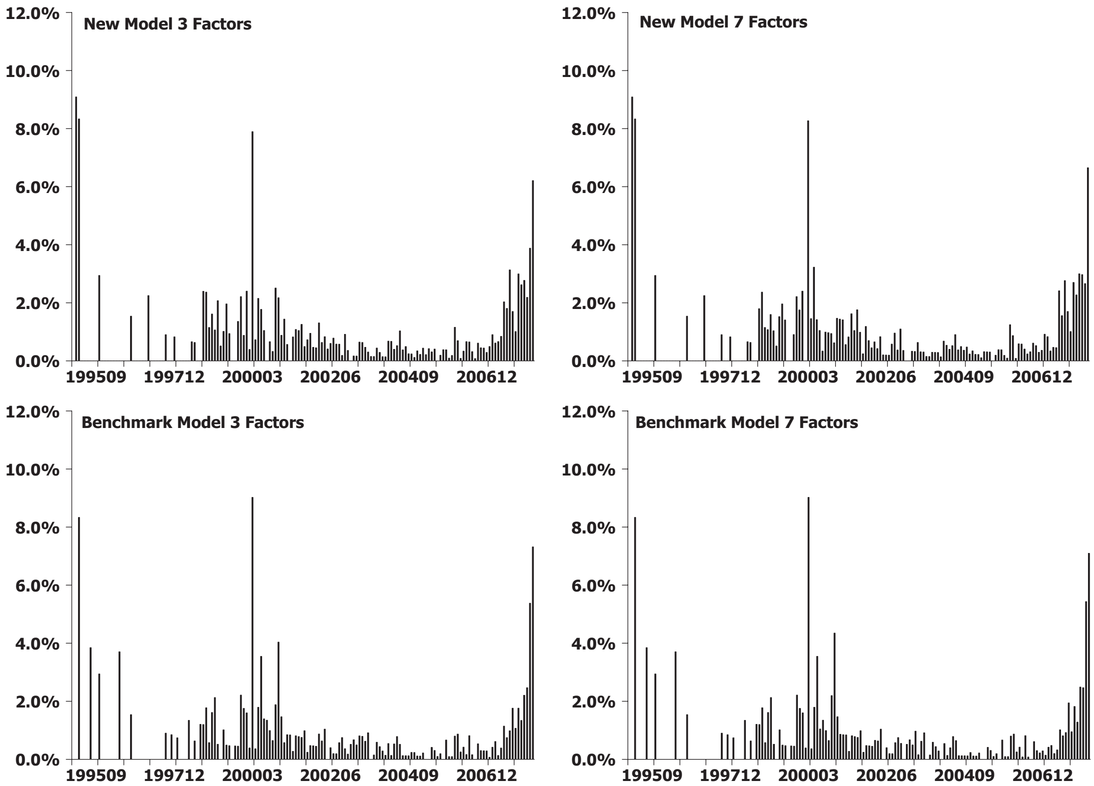


Table 1.19: Regression of Switching Frequencies on VIX/TED (Model Variation)

This table shows the results of an ordinary least squares regression of the percentage of exposure switching funds in the HF category onto the VIX index, the TED spread and a constant (intercept). Thereby, the model introduced in this study (HF New Model) is compared to a 3 factor benchmark model using the Avg-F-Test to identify switches in the factor exposure of funds (HF 3-Factor Avg-F-Test Model). The p-values (two-sided interval) are calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied.

		intercept	VIX	TED	adj R2
HF New Model	coef	−0.0057	0.0004	0.0132	0.32
	p-val	0.0124	0.0017	0.0000	
HF 3-Factor Avg-F Test Model	coef	−0.0023	0.0001	0.0155	0.22
	p-val	0.3218	0.1556	0.0065	

Table 1.20: Regression of Switching Frequencies on VIX/TED (Model Variation and Short Data Sample)

This table shows the results of an ordinary least squares regression of the percentage of exposure switching funds in the HF category onto the VIX index, the TED spread and a constant (intercept). Thereby, the model introduced in this study (HF New Model) is compared to a 3 factor benchmark model using the Avg-F-Test to identify switches in the factor exposure of funds (HF 3-Factor Avg-F-Test Model). In this analysis only the data for the time period from January 1994 until December 2005 is used. The p-values (two-sided interval) are calculated by using Newey and West (1987) heteroskedasticity and auto-correlation adjusted standard errors whereby the automated lag selection described in Newey and West (1994) is applied.

		intercept	VIX	TED	adj R2
HF New Model	coef	−0.0051	0.0004	0.0123	0.20
	p-val	0.0771	0.0054	0.0005	
HF 3-Factor Avg-F Test Model	coef	0.0010	0.0002	0.0034	0.03
	p-val	0.6046	0.0527	0.1481	

Table 1.21: Average Observations in Switching Periods

The table shows the average number of observations in each identified sub-period for hedge funds with no, one, two or three switches in each hedge fund category. The switches are identified by the structural change regression approach described in subsection 1.3.2.

		First Period	Second Period	Third Period	Fourth Period
HF	no Switch	66	0	0	0
	one Switch	44	41	0	0
	two Switches	32	30	53	0
	three Switches	35	23	40	39
CTA	no Switch	82	0	0	0
	one Switch	28	59	0	0
	two Switches	60	27	62	0
	three Switches	0	0	0	0
FoF	no Switch	66	0	0	0
	one Switch	53	28	0	0
	two Switches	40	47	39	0
	three Switches	31	20	54	32
CPO	no Switch	92	0	0	0
	one Switch	35	68	0	0
	two Switches	39	21	70	0
	three Switches	38	24	78	43
Event Driven	no Switch	67	0	0	0
	one Switch	62	33	0	0
	two Switches	25	29	32	0
	three Switches	12	40	12	80
Global Macro	no Switch	68	0	0	0
	one Switch	36	42	0	0
	two Switches	27	14	46	0
	three Switches	18	12	13	78
Multi Strategy	no Switch	68	0	0	0
	one Switch	55	31	0	0
	two Switches	41	41	39	0
	three Switches	35	24	56	31
Single Strategy	no Switch	65	0	0	0
	one Switch	48	32	0	0
	two Switches	31	32	46	0
	three Switches	21	17	76	25
Systematic	no Switch	82	0	0	0
	one Switch	33	56	0	0
	two Switches	67	30	63	0
	three Switches	0	0	0	0

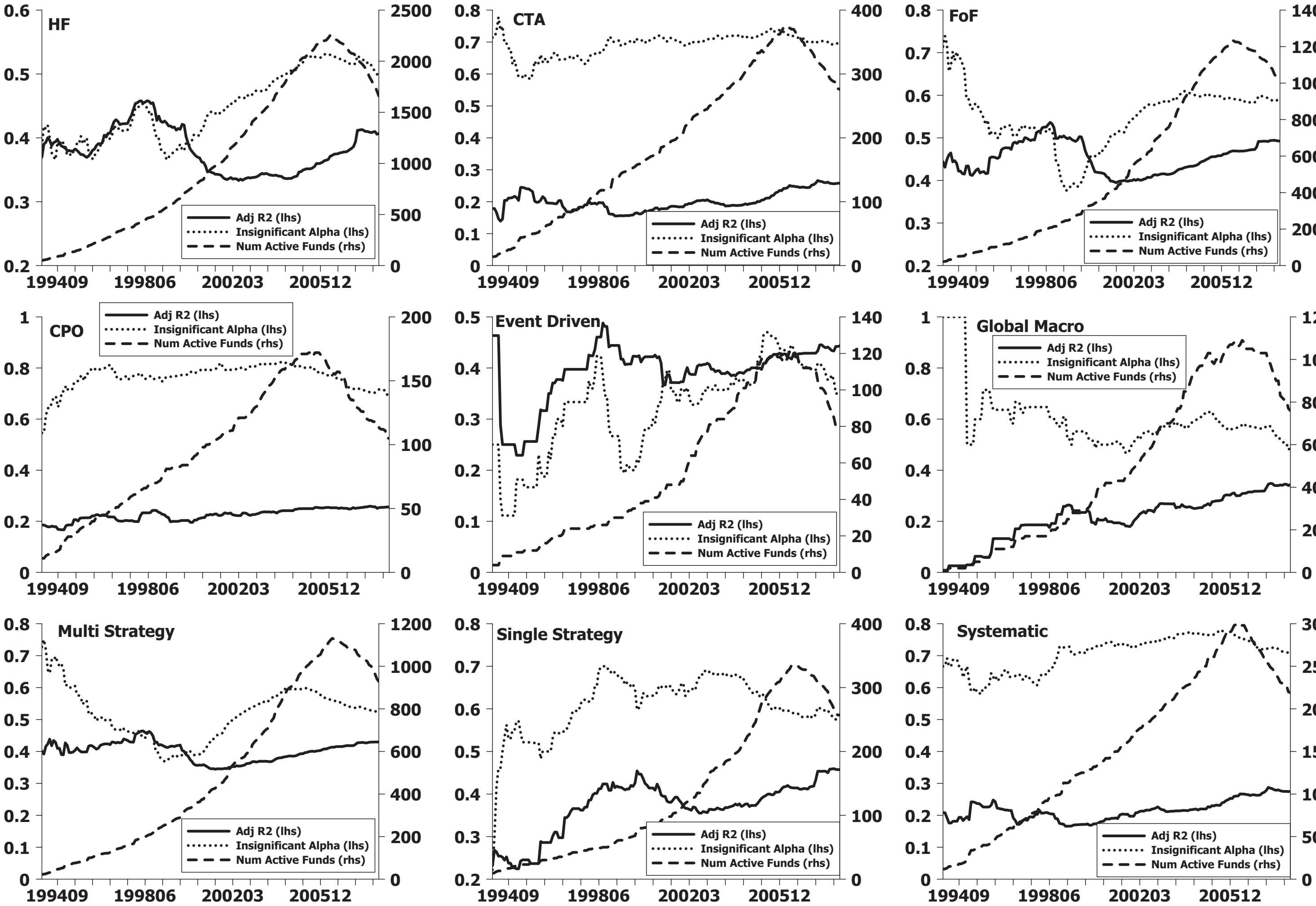
Table 1.22: Average Performance of Funds in Sub-periods

The table shows the average Sharpe ratio in each identified sub-period for hedge funds with no, one, two or three switches in each hedge fund category. The switches are identified by the structural change regression approach described in subsection 1.3.2.

		Sharpe Ratio in Subperiods			
		First Period	Second Period	Third Period	Fourth Period
HF	no Switch	0.18	0.00	0.00	0.00
	one Switch	0.53	−0.02	0.00	0.00
	two Switches	0.45	0.55	0.12	0.00
	three Switches	0.82	0.65	0.97	0.29
CTA	no Switch	0.13	0.00	0.00	0.00
	one Switch	0.40	0.15	0.00	0.00
	two Switches	0.05	0.27	0.05	0.00
	three Switches	0.00	0.00	0.00	0.00
FoF	no Switch	0.07	0.00	0.00	0.00
	one Switch	0.47	−0.36	0.00	0.00
	two Switches	0.37	0.42	−0.34	0.00
	three Switches	0.25	0.32	0.47	−0.30
CPO	no Switch	0.13	0.00	0.00	0.00
	one Switch	0.23	0.04	0.00	0.00
	two Switches	0.06	0.33	0.13	0.00
	three Switches	0.25	−0.26	0.23	−0.04
Event Driven	no Switch	0.18	0.00	0.00	0.00
	one Switch	0.62	−0.15	0.00	0.00
	two Switches	0.09	0.23	0.13	0.00
	three Switches	−0.39	1.13	−0.18	0.50
Global Macro	no Switch	0.16	0.00	0.00	0.00
	one Switch	0.32	−0.01	0.00	0.00
	two Switches	0.15	0.73	0.08	0.00
	three Switches	0.13	0.37	0.47	0.19
Multi Strategy	no Switch	0.12	0.00	0.00	0.00
	one Switch	0.48	−0.28	0.00	0.00
	two Switches	0.38	0.55	−0.29	0.00
	three Switches	0.29	0.24	0.48	−0.35
Single Strategy	no Switch	0.24	0.00	0.00	0.00
	one Switch	0.56	−0.18	0.00	0.00
	two Switches	0.42	0.54	−0.23	0.00
	three Switches	−0.16	0.20	0.30	−0.33
Systematic	no Switch	0.14	0.00	0.00	0.00
	one Switch	0.31	0.09	0.00	0.00
	two Switches	−0.02	0.27	0.02	0.00
	three Switches	0.00	0.00	0.00	0.00

Figure 1.13: Performance Appraisal Over Time (Jan 1994 - Jun 2008)

The following plots show the average adjusted R-squared, the percentage of funds with an insignificant alpha at the 95% confidence level as well as the number of active funds in the data sample for each hedge fund category at each point in time. The average adjusted R-squared measure as well as the percentage of funds with an insignificant alpha at each point in time are derived from the structural change regression approach described in subsection 1.3.2. Compared to the results in table 1.4.3, these results were found by using a constrained data sample covering the period from January 1994 to June 2008.



Chapter 2

Risk Model Performance: A Comparison Study of Risk Models Across Markets, Strategies and Risk Metrics

2.1 Introduction

Over the last decade the importance of risk management for financial institutions has increased significantly as new regulations are requiring more sophisticated risk assessment tools. Examples for this development are found in the Basel III reform package regarding banks, the Solvency 2 and the Swiss Solvency Test frameworks targeting insurance companies, as well as the UCITS IV regulations aiming at collective investment vehicles (mutual funds) in Europe. This development has been accelerated by recent market turmoils such as the sub-prime mortgage crisis of 2008 as well as the subsequent European sovereign debt crisis. The aim of this study is to provide a comprehensive comparison of the performance of different risk models at predicting the market risk of liquid financial assets and common trading strategies. Thereby, the work can be seen as an extension to the work of Kuester, Mittnik and Paolella (2006) and Angelidis, Benos and Degiannakis (2007), who have analyzed different risk models with regard to their performance of predicting the Value-at-Risk (VaR) for equity markets. Thereby, VaR is defined as follows:

$$VaR(p)_{t+h} \equiv Q_p(r_{t+h} \mid I_t) \tag{2.1}$$

In equation 2.1, Q_p is the quantile function for the return distribution of a financial asset and p defines the threshold below which $p\%$ of all observations fall. The variable r_{t+h} is defined as the return of a financial asset over the horizon $t+h$ and I_t represents the set of information available at time t. With regards to the above mentioned papers, the current study contributes the following innovations:

First, instead of only focusing on equity markets, the risk models are applied to various financial time series. In this regard, a distinction is made between static indices and common dynamic trading strategies (dynamic indices). The static indices considered are buy-and-hold investments in the following asset classes: Equities, government bonds, corporate bonds, commodities and currencies. The considered dynamic trading strategies are the following: Equity small cap premia strategy, equity value premia strategy, equity momentum strategy, commodity momentum strategy, commodity roll strategy, currency carry strategy, currency momentum strategy and currency value strategy. One aim of this study is to determine whether a certain model or certain types of models are superior at predicting risk metrics across asset classes and strategies. The results should be of interest to practitioners and researchers alike[1].

As a second contribution, in addition to analyzing the prediction of VaR for different risk models and various asset classes and strategies, this study will further investigate the ability of predicting the Expected Shortfall (ES) which is sometimes also referred to as Conditional Value-at-Risk (CVaR). The expected shortfall $ES(p)_{t+h}$ is the expected return of a financial asset

[1] The results could for example be interesting for research focused on multivariate risk modeling, where being able to choose the same risk process for every asset is an advantage over having to model different processes for the assets under consideration.

over the horizon $t+h$, conditional on the return being lower than $VaR(p)_{t+h}$:

$$ES(p)_{t+h} \equiv E[r_{t+h} \mid r_{t+h} < VaR(p)_{t+h}] \tag{2.2}$$

Thus, ES aggregates information about the entire shape of the left tail of a distribution which VaR does not and there are other arguments for choosing ES over VaR as a risk metric. First, as explained in Artzner et al. (1999), ES in comparison to VaR is coherent which refers to desirable axiomatic properties for any practical risk metric. Further, ES as a risk metric has gained increased attention from the regulatory side, whereby for example the Swiss Solvency Test requires Swiss insurance companies to manage the risks of their investment portfolio by using ES as their main risk metric. Lastly, as it is shown for example in Yamai and Yoshiba (2001), the ES risk measure is consistent with second order stochastic dominance and therefore has a direct link to the utility maximization framework. More recent studies on risk models such as for example Taylor (2008*b*) and Zhu and Galbraith (2011) have also put more emphasis on ES as a risk measure instead of solely focusing on VaR. In addition to analyzing the performance of the different risk models at predicting the VaR and ES, this study also analyzes how well the entire left tail of the distribution is being predicted by applying a goodness of fit test. By analyzing the prediction of the entire tail, conclusions may be drawn on whether certain parts of the tail are over- or under-emphasized by different models. Further, compared to the study of Kuester, Mittnik and Paolella (2006), for the VaR, ES and tail fit predictions, two forecasting horizons are analyzed in this study.

The third contribution of this study is related to the choice of the models under consideration. As a basis, a benchmark set of similar models is being analyzed as for example in Kuester, Mittnik and Paolella (2006), Angelidis, Benos and Degiannakis (2007), Taylor (2008*a*), Taylor (2008*b*) and Dimitrakopoulos, Kavussanos and Spyrou (2010). The benchmark models applied in this paper can be split into three main categories: Fully Parametric models (FP), Filtered Historical Simulation (FHS) models and Extreme Value Theory (EVT) models. In addition, this study adds recently proposed risk models such as the FP model using a generalized asymmetric student-t distribution proposed in Zhu and Galbraith (2010) and a robust EVT model following the approach of Mancini and Trojani (2011). According to the best knowledge of the author, this is the first study which includes these models in a broad comparison framework, analyzing not only the prediction of VaR, but also the prediction of ES as well as the goodness of fit of the predicted tail distribution.

This study proceeds as follows: In section 2.2, the used time series including the dynamic indices which are applied in this study to test the risk models are explained in detail. In section 2.3, the risk models as well as the statistical tests which are used to judge the performance of the risk models are introduced. Section 2.4 provides the empirical results and section 2.5 concludes by summarizing the results.

2.2 Data

2.2.1 Static Indices

A main contribution of this study is the testing of the risk models across various markets and trading strategies. Previous studies often used equity indices, single equity stocks or foreign exchange rates to test the risk models. The static indices which are covered in this study and the respective data source are the following whereby all series are quoted in daily USD excess log returns and cover the period from 31/5/1991 to 26/12/2012: For the equity market, the market factor used in Fama and French (1993) is being used and will be denoted as (EQMK)[2]. In the fixed income market, 10-year US-treasury note futures (GB10) are used. These data series are generated

[2]The data is available on Kenneth French homepage:
http://mba.tuck.dartmouth.edu/pages/faculty /ken.french/data_library.html

by using the generic future series available from Bloomberg for the front month 10-year treasury future[3]. For the corporate credit market (CRED), the Barclays US Corporate Investment Grade Total Return index is used, available via Datastream[4]. The commodity market is represented by the S&P GSCI Excess Return index (COMK), downloaded from Bloomberg[5]. To reflect the currency market, the U.S. Dollar index is chosen (CUMK), where the return series is generated by rolling front month U.S. Dollar index futures[6]. For all markets only days where the US-stock market was open are included which gives a sample of 5441 returns for every market.

2.2.2 Dynamic Indices

In addition to testing the performance of different risk models for a divers set of different liquid financial markets, this study will also introduce different liquid trading strategies for which the performance of the risk models will be tested. The motivation for including different trading strategies is twofold. First, there is a growing market of liquid trading strategies which are being offered mainly to institutional investors by investment banks[7]. This development brings the risk management of dynamic trading strategies into the focus of any involved investor. Second, there is a growing interest from the academic research side analyzing whether there are risk premia embedded in certain trading strategies and whether these premia related strategies can be used to construct more efficient portfolios (see for example

[3]The Bloomberg download codes for GB10 is TY1 Comdty. The futures are rolled 10 days prior to expiration and a ratio roll adjustment is made in order to account for the cost of rolling the futures.

[4]The Datastream code is LHCCORP. In order to obtain excess returns the daily treasury bill returns available on Kenneth French homepage are subtracted from the total return series.

[5]Bloomberg ticker SPGSCIP Index.

[6]The U.S. Dollar index represents a trade volume weighted basket of currently six currencies (EUR, JPY, GBP, CAD, SEK and CHF). The Bloomberg ticker for the generic front month U.S. Dollar index is DX1 Curncy. The same roll procedure as for the treasury note futures is applied.

[7]The available offering of trading strategies is often not disclosed. An exception is Barclays Capital which publishes their offering of trading strategies online: https://ecommerce.barcap.com/ indices/index.dxml

Asness, Moskowitz and Pedersen (2013) and Bender et al. (2010) respectively). Thus, in order to have a better understanding of the possible benefit of such strategies, the risk and the handling thereof is crucial.

The trading strategies considered in this study are motivated by academic research which finds evidence that there is a risk premium related to a specific strategy or that it has beneficial diversifying effects in a portfolio context. Also, the type of strategy should be available in practice, whereby I refer to the offering of Barclays Capital as a proxy[8]. However, the strategies do not have to be motivated by an asset pricing model and no claim is made that the following choice of strategies is complete.

Likely the best known examples of trading strategies are the small cap premium and value premium factors of Fama and French (1993) in the equity space. These factors are based on long/short trading strategies in equities, where for example the small cap premium is reflected by a strategy holding a long position in small market cap stocks against a short position in large market cap stocks. Due to the prominence of these factors they are included in this study and referred to as ESMB and EVAL respectively[9]. Further, in order to represent an equity momentum strategy as defined for example in Jegadeesh and Titman (1993), I also include the equity momentum factor (EMOM) also available on Kenneth French's homepage.

In the commodity space, the best documented trading strategies are related to momentum and backwardation strategies. Thereby, the basic idea of a backwardation strategy is to hold a long position in commodities which are in backwardation and to hold a short position in those commodities which are in contango. While it is shown in Gorton, Hayashi and Rouwenhorst (2007) that part of the excess return of commodity momentum and backwardation strategies is related to selecting commodities where inventories are low, Fuertes, Miffre and Rallis (2010) show that momentum and

[8] See https://ecommerce.barcap.com/indices/index.dxml

[9] The daily time series can be downloaded from Kenneth French's homepage. A more liquid alternative to the ESMB factor could be constructed by a long/short portfolio of Russell 2000 Index futures against S&P 500 index futures.

backwardation strategies are producing non-overlapping signals and can in fact significantly improve the risk return characteristics of a portfolio when combined. For this study a commodity momentum strategy (COMO) and a commodity backwardation strategy (COBA) have been defined using a set of daily futures data on 22 commodities from Bloomberg. The strategies apply monthly rebalancing and are derived in the appendix.

In the currency markets there has long been empirical evidence that there exists a forward premium as for example analyzed in Fama (1984). The forward premium can be extracted by following a carry trade strategy where an investor borrows in a low interest rate currency and invests the proceeds in a higher interest rate currency. The carry trade has been studied extensively in the academic literature and there is evidence that the carry trade premium is at least in part compensation for taking on specific systematic risk as shown in Christiansen, Ranaldo and Söderlind (2011) and Lustig, Roussanov and Verdelhan (2011). In addition, there exist other strategies in the foreign exchange market. The most prominent ones are momentum and value strategies as described for example in the study of Pojarliev and Levich (2008). Thereby, a value strategy takes a long position in currencies which are undervalued based on fundamental data and takes a short position in overvalued currencies. In this study, the risk models will be tested on three currency strategies: A currency carry strategy (CUCA), a momentum strategy (CUMO) and a value strategy (CUVA). The currency strategies are based on a set of 12 exchange rates against the USD with monthly rebalancing. The construction of the strategies is explained in the appendix. A statistical summary of all considered indices, showing the mean (Mean), standard deviation (Std), skewness (Skew) and kurtosis (Kurt) of the return series of every index is found in table 2.1 [10].

[10] Calculated with daily data.

Table 2.1: Statistics of Indices (1-Day Returns)

The table shows the mean (Mean), standard deviation (Std), skewness (Skew) and kurtosis (Kurt) of the index series used in this study. All metrics are calculated by using daily log-returns for the period from 31/5/1991 to 26/12/2012.

	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA
Mean	0.021%	0.005%	0.015%	0.024%	0.018%	0.016%	0.001%	0.042%	0.053%	−0.007%	0.022%	0.013%	0.020%
Std	1.178%	0.578%	0.604%	0.911%	0.395%	0.334%	1.347%	1.383%	1.108%	0.550%	0.541%	0.641%	0.457%
Skew	−0.28	−0.33	0.03	−1.10	−0.13	−0.30	−0.26	0.00	−0.13	−0.04	−0.37	−0.18	−0.05
Kurt	11.07	7.44	9.23	14.93	5.99	5.57	6.34	5.25	4.93	4.66	8.55	10.11	9.46

2.3 Methodology

2.3.1 Risk Models

The risk models analyzed in this study can be distinguished into four categories: Fully Parametric (FP) models, Filtered Historical Simulation (FHS) models, Extreme Value Theory (EVT) models and Robust Extreme Value Theory (REVT) models. The following subsections give a brief introduction to the motivation and mechanics of the chosen model types.

Fully Parametric (FP) Models

The models considered in this subsection are referred to as Fully Parametric (FP) models since the assumptions about the underlying process of the financial return series is fully parametrized. This means in particular, that an explicit assumption about the distribution of the conditional returns is made. All models considered in this study are motivated by the stylized fact of volatility clustering and long term volatility mean reversion which is present in most financial time series. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model introduced by Bollerslev (1986) is able to reflect these stylized facts and has become, with its various extensions and enhancements, a standard for modeling the volatility of financial time series. In its basic form, the process for the variance σ_t^2 of the GARCH(p,q) model is defined as follows:

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2 \tag{2.3}$$

$$r_t = \mu_t + \epsilon_t = \mu_t + \sigma_t z_t \tag{2.4}$$

$$z_t \overset{iid}{\sim} d(0, 1, \Lambda) \tag{2.5}$$

In the above equation, r_t refers to observed returns of the financial time series and α_j, β_k, ω and μ_t are the parameters of the model which need to be estimated. The iid random variable z_t has a mean equal to 0, a standard deviation of 1 and may depend on a parameter set Λ, which needs to be estimated as well. In addition, for the drift term μ_t a process may be formulated whereby it is common that an ARMA(l,m) process as defined below is assumed.

$$\mu_t = \upsilon + \sum_{j=1}^{l} \gamma_j r_{t-j} + \sum_{k=1}^{m} \delta_k \epsilon_{t-k} \tag{2.6}$$

The parameters υ, γ_j and δ_k are estimated together with the parameters for the variance process. Given the general specification of the model in the above equations, the choice of the models analyzed in this study need to be curtailed. The first confinement is done with respect to the number of lags which are included in the model. As it is shown for example in Hansen and Lunde (2005) or Dimitrakopoulos, Kavussanos and Spyrou (2010), GARCH models which incorporate more than one lag for p and q are generally not superior to a GARCH(1,1) specification. Also, most studies focused on risk modeling use this specification and I therefore also choose a GARCH(1,1) specification. Next, the ARMA process specification needs to be determined. Most previous studies use either an AR(1) specification where $l = 1$ and $m = 0$ in equation 2.6, an ARMA(1,1) specification where $l = 1$ and $m = 1$ or a specification where μ is assumed to be constant[11]. In this study an AR(1) specification is chosen for the return process in order to account for the possibility of first order autocorrelation in the time series while keeping the model parsimonious for estimation purposes. Thus, the basic model specification in this study is set to an AR(1)-GARCH(1,1) model.

While the process specification defined above is very similar for many

[11] For example McNeil and Frey (2000), Kuester, Mittnik and Paolella (2006) and Mancini and Trojani (2011) use an AR(1) specification, while Zhu and Galbraith (2011) and Angelidis, Benos and Degiannakis (2007) assume μ to be constant. Dimitrakopoulos, Kavussanos and Spyrou (2010) use an ARMA(1,1) specification as well as an adaptive approach, where the optimal lag is found by minimizing the Schwartz Bayesian Information Criterion. However, they conclude that the adaptive approach does not yield superior out-of-sample results compared to the single lag specification.

studies in this field, the key parameter driving the performance of the risk models is the distributional assumption of the innovations z_t. While in the original setting of Bollerslev (1986), z_t is assumed to be normally distributed, it was soon realized that this assumption is flawed when applied in the context of financial time series. Therefore, Bollerslev (1987) proposed a GARCH setup where the innovations follow a t-distribution and Hansen (1994) introduced the use of a skewed-t-distribution in the same context. Different studies such as Mittnik and Paolella (2000), Giot and Laurent (2004), and Alberg, Shalit and Yosef (2008) have shown that the use of a GARCH setup with a skewed-t-distribution performs very well at predicting different risk metrics, whereby the focus in these studies was laid on the equity and currency markets. Based on these findings and the setup of previous comparison studies such as Kuester, Mittnik and Paolella (2006), three benchmark FP model specifications for the innovations are chosen. Namely normal, t-distributed and skewed-t-distributed innovations. The processes for the three benchmark FP models are all defined as an AR(1)-GARCH(1,1) process and the three benchmark models will be referred to as FP-QML, FP-T, and FP-ST respectively. In order to accommodate for recent developments one additional FP models will be analyzed in this study. This is the FP model with a generalized asymmetric t-distribution (FP-AST) which is analyzed in Zhu and Galbraith (2011) in the context of forecasting the ES for the Standard & Poor's (S&P) 500 equity index.

All parameters in the FP models are calculated by using maximum likelihood estimation. The VaR and ES estimates for a one period horizon can be calculated analytically for all FP models considered in this study[12]. For multi-horizon forecasts, Monte Carlo simulations are conducted whereby in this study 10000 return paths w for each model and each date are generated

[12] A good reference for the analytical formulas for VaR and ES of the GARCH(1,1) model with t and skewed-t distributed innovations is Christoffersen (2012). For the asymmetric t-distribution the interested reader is referred to Zhu and Galbraith (2010). The original Matlab code for estimating the generalized asymmetric student-t distribution is being provided by the authors under the following link: http://www.runmycode.org/CompanionSite/site.do?siteId=153. The code was adapted in order to fit the model setup used in this study.

to forecast the return distribution. The process for generating a simulation path is iterative and can be described as follows: First, a random variable z_{t+1}^w as in equation 2.5 is drawn from the distribution defined for the FP model under consideration. Second, having estimated the parameters of the GARCH and ARMA models in equations 2.3 and 2.6 respectively, a forecast for the variance $\tilde{\sigma}_{t+1}^2$ and the drift $\tilde{\mu}_{t+1}$ are calculated and together with the drawn z_{t+1}^w random variable using equation 2.4 a realization of a future return r_{t+1}^w can be generated. In order to find the next return forecast r_{t+2}^w, the same procedure is repeated, whereby now the new variance and drift forecasts ($\tilde{\sigma}_{t+2}^{2w}$ and $\tilde{\mu}_{t+2}^w$) are used. This procedure is repeated for $w = 1, 2, ..., 10000$.

Filtered Historical Simulation (FHS) Models

Filtered Historical Simulation (FHS) has been shown to perform strongly in a risk management context for example in Barone-Adesi, Giannopoulos and Vosper (2002), Pritsker (2006) and Kuester, Mittnik and Paolella (2006)[13].

The concept of (FHS) is linked to the FP modeling approach by using the concept of a dynamic variance as portrayed in equation 2.3. However, the FHS models do not make any assumptions about the distributional properties of the innovations z_t, but rely on actual historical realizations to model the innovations. The process for obtaining the return distribution in a FHS framework can be split into three steps. In a first step, the parameters of a GARCH type model, in our case an AR(1)-GARCH(1,1), are estimated in the same way as it is done for the FP models. Then, in a second step, all historical returns used to estimate the model parameters are taken to find the corresponding historical realizations of the innovations by solving equation 2.4 for z_t for every point of the estimation window. Thus, if an estimation window of $G = 1000$ returns was used, a set of 1000 historical innovations $\hat{Z}_G$ is obtained. In a third step, the forecast of the return distribution can be obtained by using the current forecasts for the variance and the drift, and

[13] Barone-Adesi, Giannopoulos and Vosper (2002) as well as Pritsker (2006) also give a detailed introduction into the mechanics of the FHS approach.

subsequently plugging-in every $\hat{z}_g$ into equation 2.4, which gives $G = 1000$ one-step ahead realizations of r_{t+1}. From the distribution of these realizations it is straight forward to obtain VaR, ES and any other desired risk metric. If multi-horizon forecasts are required, the same simulation procedure can be used as described for the FP models where instead of drawing randomly from the defined parametric distribution, the random draws with replacement are made from the set of historical innovations $\hat{Z}_G$.

When estimating an FHS model the question arises which assumption should be made about the distribution of the innovations in the filtering of the volatility process. As it has been shown in Bollerslev and Wooldridge (1992), using the assumption of normally distributed innovations and estimating the parameters with maximum likelihood gives parameter estimates which converge to the true parameter estimates even when the assumption of normality is flawed[14]. In this case the estimation is referred to as quasi-maximum likelihood estimation (QMLE). Thus, using QMLE in the context of FHS seems to be an obvious choice as we do not make any assumption about the distribution of the innovations. However, in a finite sample setting the use of QMLE may lead to miss-specified parameters as the estimates only converge to the true values asymptotically. For practical purposes this problem can be tackled by assuming that the innovations follow a more flexible distribution, such as the skewed-t distribution. As the results of Kuester, Mittnik and Paolella (2006) point out, using the skewed-t distribution assumption in the volatility filtering process may influence the performance of the models when compared to a QMLE approach. Therefore, in this study three versions of the FHS model will be used. The first model referred to as FHS-QML uses the QMLE approach to estimate the parameters of the GARCH process, while the second model, referred to as FHS-ST, uses the assumption of skewed-t innovations in the estimation process. As a third variation of the FHS approach, this study introduces the FHS-AST model

[14] Assuming that the processes of the variance and the mean are correctly specified.

which applies the assumption of generalized asymmetric t-distributed innovations in the estimation procedure.

Extreme Value Theory (EVT) Models

The third class of models analyzed in this study are Extreme Value Theory (EVT) models whereby the focus will be on the model type introduced in McNeil and Frey (2000), which combine a filtering process similar to the FHS models with an estimation of the tail of the distribution using an extreme value theory approach. As shown in Kuester, Mittnik and Paolella (2006) and Angelidis, Benos and Degiannakis (2007), this type of model performs very well at predicting VaR, especially at low values for p (e.g 1%, 2.5%). Unconditional EVT models which leave out the filtering process are not considered in this study[15].

The EVT approach applied in this study can be split into two phases. In the first phase, the same procedure is being run as in the FHS approach, leading to a set of historical innovations $\hat{Z}_G$. In the second phase, instead of using directly the distribution of the historical innovations in $\hat{Z}_G$, a generalized Pareto distribution (GPD) is fit to the tail of the historical innovations where the number of innovations included for the estimation is being determined by the threshold level u. The motivation for fitting a GPD to the tail of the innovation comes from a result in extreme value theory, which states that for a large class of distributions the observations in the tail beyond the threshold level u, here denoted as y, converge to a GPD where the tail index

[15] As is shown in Kuester, Mittnik and Paolella (2006) unconditional approaches perform worse at predicting VaR compared to all conditional approaches considered in this study. Results comparing the performance of unconditional EVT approaches can be found in Bekiros and Georgoutsos (2005) and Brooks et al. (2005).

ξ and the scale parameter ϑ need to be estimated[16].

$$GPD(y) = \begin{cases} 1 - exp(-y/\vartheta) & \text{if } \xi = 0 \\ 1 - (1 + \xi y/\vartheta)^{-1/\xi} & \text{if } \xi \neq 0 \end{cases} \tag{2.7}$$

In this study, the estimation of the GPD in the tail is done my maximum likelihood estimation[17]. The threshold level u in this study is chosen such that 10% of the observations in the tail are used to fit the tail to the GPD. Since in this study $G = 1000$ observations are used for fitting the GARCH process, $K = 100$ observations are used to fit the GPD to the tail. This choice follows the findings of McNeil and Frey (2000), who show in a simulation study with a sample of 1000 observations that the mean squared error of the GPD maximum likelihood estimator is the lowest when setting K to a value of about 100. Once the tail index ξ and the scale parameter ϑ are estimated, VaR and ES for the one period ahead horizon can be calculated analytically as follows, shown for example in Kuester, Mittnik and Paolella (2006) respectively McNeil and Frey (2000):

$$VaR(p)_{t+1} = -\left(u + \frac{\vartheta}{\xi}\left(\left(\frac{p}{K/G}\right)^{-\xi} - 1\right)\right)\tilde{\sigma}_{t+1} + \tilde{\mu}_{t+1} \tag{2.8}$$

$$ES(p)_{t+1} = -Q_Z\left(\frac{1}{1-\xi} + \frac{\vartheta - \xi u}{(1-\xi)\,Q_z}\right)\tilde{\sigma}_{t+1} + \tilde{\mu}_{t+1} \tag{2.9}$$

$$where$$

$$Q_Z = \left(u + \frac{\vartheta}{\xi}\left(\left(\frac{p}{K/G}\right)^{-\xi} - 1\right)\right)$$

In the above equations, G is the number of observations in the set $\hat{Z}_G$ while $\tilde{\sigma}_{t+1}$ and $\tilde{\mu}_{t+1}$ are the volatility and drift forecasts respectively for one

[16]See McNeil and Frey (2000) and the references therein for a detailed examination of the approach.

[17]The Matlab function gpfit is being used.

period. In order to be able to generate multi-horizon forecasts, a GPD is also fitted to the upper tail of the distribution where again $K = 100$ is chosen for the threshold.

Generating the return distribution for horizons greater than one proves to be a bit more sophisticated. The procedure chosen in this study is similar to the one applied in McNeil and Frey (2000). The approach is based on the procedure used for the FHS models with the difference that the random draw from the set of innovations $\hat{Z}_G$ is replaced whenever it belongs to the K highest or lowest returns available in $\hat{Z}_G$. If the draw belongs to the K lowest returns, it is being replaced by a random draw from the GPD which has been fit to the left tail of the distribution of historical innovations. Whenever a draw belongs to the K highest returns, it is replaced by a random draw from the GDP which has been fitted to the right tail of the distribution of historical innovations. If a draw belongs neither to the highest or lowest K returns in $\hat{Z}_G$, it is unadjusted as it is the case for the FHS models.

Similar to the case of the FHS models, the question arises whether to use QMLE estimation or to use a specific assumption about the distribution of the innovation in the estimation process of the GARCH parameters. Following the argumentation for the FHS models, in this study I will use three filtering processes for the EVT models. The first one, referred to as EVT-QML, uses QMLE for the parameter estimation and the second one, referred to as EVT-ST, uses the assumption of skewed-t distributed innovations in the estimation process. The third variation EVT-AST introduced in this study, applies the assumption of generalized asymmetric t-distributed innovation in the estimation procedure.

Robust Extreme Value Theory (REVT) Models

The fourth class of models considered in this study are an extension of the EVT approach and was proposed by Mancini and Trojani (2011). Their idea consists of applying robust estimators for the parameters used in the EVT-

QML model. The motivation for using robust estimators for the parameters in the EVT-QML setup is driven by the findings in Mancini, Ronchetti and Trojani (2005) which show that even slight misspecifications of the GARCH-dynamics may lead to parameter estimates which imply dynamics that are materially different form the true dynamics. Similarly, when the residuals in the tail which are fitted to a GPD in the EVT approach are from a slightly different distribution, the pseudo-true values and asymptotic variances of the tail-estimator may exhibit large variations (Mancini and Trojani (2011)). The robust estimators used by Mancini and Trojani (2011) belong to a class of so called M-estimators which feature bounded influence functions. Loosely speaking, the influence function of an estimator describes the effect which outliers in the data set have on the parameter estimates. Estimators with bounded influence functions ensure that the potential damaging effects of extreme data points are capped by down-weighting those observations in the estimation process. In the appendix a brief intuitive introduction is given to the idea of robust estimation procedures. The actual estimators used by Mancini and Trojani (2011) are based on their estimator developed in Mancini, Ronchetti and Trojani (2005) for the estimation of the GARCH dynamics and the robust estimator for the GPD as applied by Dupuis (1999). Both robust estimators used in this study, for the GARCH dynamics as well as the GPD parameter estimations, are defined by the specification shown below:

$$E\left[\psi_c\left(s\left(Y,\theta\right)\right)\right] = 0 \tag{2.10}$$

$$where$$

$$\psi_c = A\left(\theta\right)\left(s(Y,\theta) - \tau\left(Y,\theta\right)\right) w\left(Y,\theta,c\right) \tag{2.11}$$

$$w\left(Y,\theta,c\right) := min\left(1, c\left\|A\left(\theta\right)s\left(Y,\theta\right) - \tau\left(Y,\theta\right)\right\|^{-1}\right) \tag{2.12}$$

Thereby, $s\left(Y,\theta\right)$ is the score function defined as the gradient of the maximum likelihood estimator with parameters θ applied on the set of observations Y

and $w(Y, \theta, c)$ is the weighting function whereby the constant c determines the degree of robustness. The matrix $A(\theta)$ and the vector $\tau(Y, \theta)$ are implicitly defined as the solution to the following equations:

$$E\left[\psi_c(s(Y, \theta))\,\psi_c(s(Y, \theta))^{\intercal}\right] = I$$

$$E\left[\psi_c(s(Y, \theta))\right] = 0$$

As there is a trade-off between robustness and efficiency of the estimator I follow the results of Mancini, Ronchetti and Trojani (2005) and Dupuis (1999) and choose c to be equal to 11 and 4 for the robust GARCH estimation and the robust GPD estimation respectively[18].

In this study a slightly altered version of the model applied in Mancini and Trojani (2011) is used and will be referred to as REVT-ROB. The main difference is the neglected leverage term used in the original model. Also, the model is applied to a rolling window of 1000 observations compared to a rolling window of 2000 observations in Mancini and Trojani (2011). Further, in order to investigate whether the robust EVT estimation is able to generally improve the empirical forecasting results, three additional models are specified in this study and will be referred to as REVT-QML, REVT-ST and REVT-AST. These models can be regarded as enhanced EVT models where the GARCH dynamics are the same as in the EVT-QML, EVT-ST and EVT-AST models, but where the fitting of the GPD is done by the robust estimator described above.

[18] I would like to thank Professor Mancini who has provided the code for the robust GARCH estimation used in Mancini and Trojani (2011). The code has been altered to match the model specification used in this study. The code for the robust GPD estimation is based on the FORTRAN code used in Dupuis (1999) which has been translated into Matlab code by Igor Moiseev. The translated code is available under the following link: https://github.com/moiseevigor/obre/tree/master/matlab

Models Not Considered

There are numerous other types of risk models which have been proposed in the literature and which are not discussed in this study. A prominent class of models are unconditional models which do not take into account volatility clusters which are often present in financial time series. As shown in Kuester, Mittnik and Paolella (2006), they tend to significantly underperform conditional models and are therefore not further investigated in this study. Quantile regression models as analyzed and further developed for example in Taylor (2008*a*) and Taylor (2008*b*) are also not considered in this study. The reasons for this are twofold. First, although the empirical result for example in Taylor (2008*a*) indicate that the more sophisticated types of quantile models are competitive when compared to the GARCH-type conditional models, there is no clear indication for a significant outperformance. Second, in order to avoid an overcrowding of the current study, quantile regression based methods are excluded as they are based on an entirely different technical approach compared to the methods discussed in this study. For similar reasons mixture GARCH models as proposed for example in Haas, Mittnik and Paolella (2004) are not considered in this study. As documented in Kuester, Mittnik and Paolella (2006), these types of models have been inferior to certain EVT and FHS type models in their test setting. Further, there is a very wide range of possible model specification and it is not clear which ones should be considered. Some guidance with regards to promising model specifications for mixed GARCH models can be found in the recent study of Broda et al. (2013). Further, other GARCH process specifications such as for example the long memory GARCH model (HYGARCH) of Davidson (2004) are not considered in this study but may be a promising field for future research. Also, the less prominent risk modeling approach of using Markov-switching models as proposed for example in Elliott and Miao (2009) is not considered in this study. Finally, similar to Kuester, Mittnik and Paolella (2006), methods which use intraday data or implied volatility data are not treated in this study due to the lack of appropriate data given the chosen indices.

2.3.2 Statistical Tests for Comparing the Risk Models

In this study a set of different statistical tests is being used to judge the performance of the different risk models. For testing the accuracy of predicting VaR, the test of unconditional coverage, the test of independence as well as the test of conditional coverage are being considered. For testing the accuracy of the ES prediction, a mean zero test similar to the one proposed in McNeil and Frey (2000) is being used. For testing the goodness-of-fit of the tail distribution predictions, a Pearson Chi-Square test is being applied.

Test of Unconditional Coverage

The test of unconditional coverage developed in Kupiec (1995), tests whether the number of realized violations of VaR ($r_{t+h} < VaR(p)_{t+h}$) is significantly different from the expected number of violations given by p. The test is formulated as the likelihood-ratio statistic shown below which follows a chi-square distribution with one degree of freedom χ_1^2. Under the null hypothesis, the realized number of violations correspond to the predicted number of violations. The number of observation in the time period is denoted by T and the number of violations registered during this period is denoted by V.

$$LR_{uc} = 2\left(ln\left(\left(1-\frac{V}{T}\right)^{T-V}\left(\frac{V}{T}\right)^{V}\right) - ln\left((1-p)^{T-V}p^{V}\right)\right) \sim \chi_1^2 \tag{2.13}$$

Test of Independence

The test of independence suggested by Christoffersen (1998), tests whether the sequence of VaR violations is independent over time. Independence of VaR violations is a crucial feature for practical risk management purposes of any risk model. If violations occur in clusters, an available risk budget for a given time period based on a VaR forecast is obsolete, as the probability of depleting the given risk budget over the given time period can not be predicted by the flawed model[19]. The test is conducted by splitting the observations in the sample T into four groups whereby the first observation is excluded as no conditional observation is available: T_{00} is the number of non-violations followed by a non-violation, $T_{01}$is the number of non-violations followed by a violation, $T_{11}$is the number of violations followed by another violations and T_{10} are the number of violations followed by a non-violation. The test is again a likelihood ratio statistic where under the null hypothesis the violations of VaR are independent over time.

$$LR_{ind} = 2\left(ln\left((1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T11}\right) - ln\left((1-\pi_2)^{T_{00}+T_{01}}\pi_2^{T_{01}+T_{11}}\right)\right) \sim \chi_1^2 \tag{2.14}$$

$$\text{where} \quad \pi_{01} = \frac{T_{01}}{T_{00}+T_{01}},\ \pi_{11} = \frac{T_{11}}{T_{10}+T_{11}},\ \pi_{00} = 1-\pi_{01},\ \pi_{10} = 1-\pi_{11},\ \pi_2 = \frac{T_{01}+T_{11}}{T-1}$$

Test of Conditional Coverage

In order to test simultaneously the null hypothesis that the predicted number of violations is correct and that the violations are independent over time, Christoffersen (1998) proposed the test of conditional coverage. He shows that when conditioning on the first observation, the likelihood ratio statistic of the test of conditional coverage is simply the sum of the likelihood ratio statistics of the test of independence and the test of unconditional coverage.

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2 \tag{2.15}$$

[19]Although the model may predict correctly the expected number of violations on average, as verified by the test of unconditional coverage.

Mean Zero Test

To test the performance of the different models with regards to their ability to predict the ES, I conduct a mean zero test as proposed in McNeil and Frey (2000). The idea is to test whether, on average, the realized return conditional on an occurred violation is different from the estimated ES. This is done by testing whether the mean of the difference between the realized return r_{t+h} and the predicted expected shortfall $ES(p)_{t+h}$ is zero. This difference is define as the exceedance residual er_{t+h} whereby an adjustment is made for the time-varying volatility of the forecasts.

$$er_{t+h} = \frac{r_{t+1} - ES(p)_{t+h}}{\tilde{\sigma}_{t+h}}, \text{ where } \{er_{t+h} : t \in T,\ r_{t+h} < VaR(p)_{t+h}\} \tag{2.16}$$

The null hypothesis of the test states that the exceedance residuals have a mean equal to 0. In order to test this hypothesis, an ordinary t-test can be applied or, when no assumption about the actual distribution of the underlyings is made, a bootstrap test can be conducted. In this study a two sided bootstrap-test is applied to the exceedance residuals. McNeil and Frey (2000) apply a one-sided bootstrap test against the alternative hypothesis that the residuals have a mean smaller than zero. When choosing a one-sided test, the analysis only detects an underestimation of ES. However, since having too conservative risk metrics can also bear significant opportunity costs, in this study a possible overestimation of ES should also be revealed. Therefore, a two-sided mean zero test is chosen instead of a one-sided test.

Goodness of Fit Test

Instead of only focusing on risk measures like VaR and ES it may also be of interest how well the different risk models predict the entire left tail of the return distribution. In this study the focus is on conditional models where the parameters are re-estimated for every new observation on a rolling window basis. This approach brings the advantage of more accurate predictions of the conditional distribution, however a drawback is the variability of the conditional distribution. Therefore, testing the goodness-of-fit of the predicted tail cannot be conducted directly, since the shape of the conditional distribution may change from observation to observation. A way to circumvent this problem is to apply a transformation of the realized returns in the tail, such that the resulting distribution can be tested. The approach taken in this study is to transform the realized returns in the tail into a discrete uniform distribution. For the purpose of this test, the tail is defined as the 5% quantile of the return distribution. The transformation and testing procedure works as follows:

1. Quantile counter variables C_f for $f = 1, 2, ...F$ are defined where $F = 5$ in this study. The purpose of these counter variables is to count the number of returns in the tail which fall into a specific quantile according to the predicted conditional return distribution. Thereby, C_1 counts the number of returns which are in the]-Inf, 1%] range quantile, C_2 counts the number of returns in the]1%, 2%] range quantile etc. The initial value for all C_f is set equal to zero.

2. For each return $r_{t+h} < VaR(5\%)_{t+h}$ the quantile according to the predicted conditional return distribution is determined and the counter variable which covers the quantile range for the specific return is increased by +1. For example, if according to the conditional return distribution the quantile of a realized return is determined to be 3.5%, the counter variable C_4 is increased by +1.

3. If the risk model on average predicts the tail distribution correctly, the number of returns in the tail falling into the different quantiles should be equal. Therefore the counter variables C_f should represent a discrete uniform distribution. The null hypothesis that the counter variables C_f are indeed uniformly distributed

can be tested by a Pearson Chi-Squared test as shown below. The variable t_{tail} is the total number of returns in the tail satisfying $r_{t+h} < VaR(5\%)_{t+h}$.

$$\sum_{f=1}^{F} \frac{\left(C_f - \frac{t_{tail}}{F}\right)}{\frac{t_{tail}}{F}} \sim \chi^2_{F-1} \tag{2.17}$$

The goodness of fit test for the prediction of the tail distribution could also be conducted by transforming the returns in the tail to a continuous uniform or normal distribution as proposed for example in Christoffersen (2012). However, due to the limited data history the number of observations in the tail shrink considerably once longer horizon forecasts are made. In this context the data points may be too sparse in order to conduct reliable goodness of fit tests for the continuous distributions. The proposed Pearson Chi-Square test however is still reliable for the available data points in the context of longer horizon forecasts.

2.4 Empirical Results

In this section the tests described in section 2.3.2 are used to test the out-of-sample performance of the models described in section 2.3.1. Thereby, the models are applied to the different static and dynamic indices described in chapter 2.2. For all tests, two forecasting horizons are considered, namely 1-day and 5-days. The parameters for the models are re-estimated for every day in the sample whereby a rolling window of 1000 returns is used. This results in 4441 test observations for the 1-day horizon and 888 test observations for the 5-day horizon.

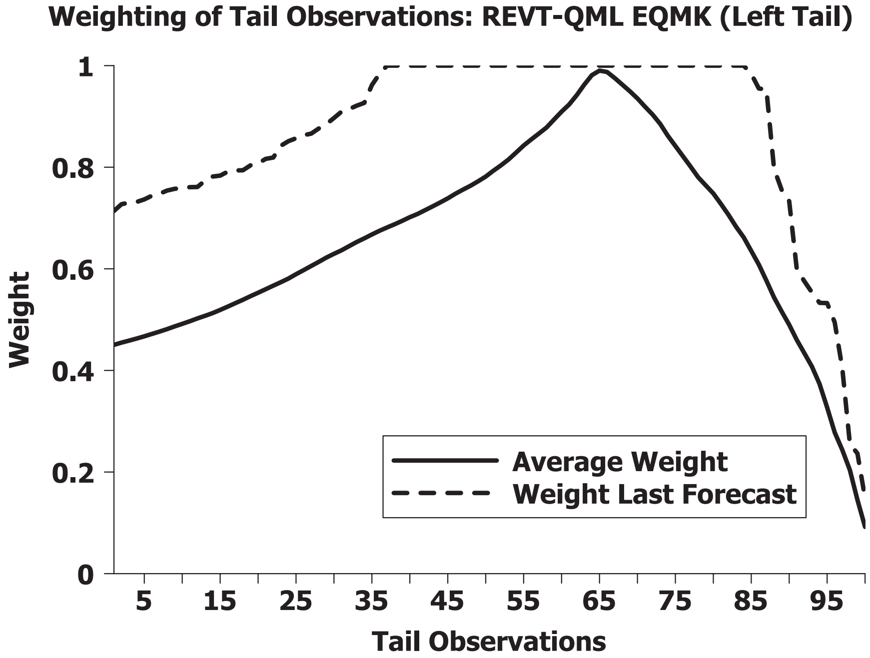
2.4.1 Testing the Performance of Value-at-Risk (VaR) Predictions

The first subsection of this chapter holds the results of the performance tests of the different risk models with regard to predicting the Value-at-Risk (VaR) risk metric. The first test under consideration is the test of unconditional coverage, which tests whether the predicted number of VaR violations correspond to the actual observed number of VaR exceedances. Table 2.2 shows the p-values for the null hypothesis of the unconditional coverage test for the different models and indices for the 1-day forecasting horizon and table 2.3 holds the results for the 5-day forecasting horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1% for every test. In order to distinguish under- and over-estimations of risk, the p-values for which the VaR forecasts are too conservative (over-estimation) carry a negative sign. In order to summarize the results of the p-value tables, I follow a similar approach as applied in Taylor (2008*a*) respectively Taylor (2008*b*), by counting the number of p-values for which the null hypothesis is rejected at the 5% significance level for each model. The summarized results for the unconditional coverage test are shown in table 2.4, whereby for each threshold level and forecasting horizon over- and under-estimations are separated such that more details about the performance of the different risk models are revealed.

The results for the 1-day forecasting horizon in table 2.4 show that when considering the overall performance (column: *All 1d Total*), the best performing models with only 4 rejections are the FP-ST and the FP-AST model, followed by the REVT-ROB and EVT-AST model with 6 respectively 8 rejections. The weakest performing model is the simplest model FP-QML which confirms the results of prior studies. When analyzing the 5% and 1% threshold levels separately, an interesting observation is made with respect to the robust models (REVT-QML, REVT-ST, REVT-AST and REVT-ROB) which show the fewest rejections of all models at the 5% level (column: *5% 1d Total*) but are the only models which show significant under- *and* over-estimation of risk at the 1% threshold level (column: *1% 1d Over*). These results give an indication, that the risk in the far tail (1% threshold level) is not captured as well by the robust methods when compared to the more sophisticated full parametric models FP-ST and FP-AST. A reason for this outcome may be related to the weights assigned to the observations in the extremes of the tail which are generally lower than the weights given to the bulk of the observations in the tail. Such a weighting scheme is disadvantageous during distressed markets as experienced in the period under consideration. The weighting scheme is illustrated in figure 2.1, where the average weight assigned to the $K = 100$ observations in the tail are plotted by the solid line for the REVT-QML model applied on the EQMK index. The dashed line also shows the specific weighting applied for the last forecast using the REVT-QML model on the EQMK index for illustrative purposes. Nevertheless, overall the REVT-ROB model is the second best performing model at the 1-day forecasting horizon for the unconditional coverage ratio test.

Figure 2.1: Weighting of Tail Observations

The solid line in the graph depicts the average weight assigned to the $K = 100$ observations in the tail for the REVT-QML model applied on the EQMK index for the 4441 1-day forecasts. The observations are sorted from the smallest to the largest absolute value from left to right. For illustrative purposes, the dashed line also shows the weighting applied for the last forecast using the REVT-QML model on the EQMK index. This correspond to the weights used to obtain the 4441th 1-day forecast.



When analyzing the forecasting performance with respect to the different indices for which the number of rejections are listed in the bottom row of table 2.2, it can be seen that there are indices for which the models under consideration have more difficulties at forecasting VaR than for others. The most challenging indices seem to be EVAL with 23 rejections and COMK with 22 rejections.

When moving to the results for the 5-day forecasting horizon in table 2.4, an interesting finding is that all models perform better or equal compared to the 1-day horizon (column: *All 5d Total*). The best performing models with one rejection are the robust models REVT-QML and REVT-ROB as well as the FP-AST model. The worst performing model is surprisingly the FP-ST model which belonged to the best performing models at the 1-day forecasting horizon. Interestingly, the weak performance is solely the result of a strong over-estimation of risk at the 5% threshold level. Analyzing the forecasting performance with respect to the different indices at the 5-day horizon reveals, that problematic indices with many rejections at the 1-day horizon are not posing the same challenge when analyzing the 5-day horizon. The index with the highest number of rejections (10) is EVAL.

Adding together the results of the 1-day forecasting horizon and the 5-day forecasting horizon in table 2.4, the best performing overall models with regards to the unconditional coverage test is the FP-AST model with 5 rejections of the null hypothesis at the 5% significance level, followed by the REVT-ROB model with 7 rejections (column: *All Total*). Considering only under-estimations of risk, the best performing models overall are the REVT-ROB model with 3 rejections, followed by the FP-ST and FP-AST models with 4 rejections each (column: *All Under*). In order to have an alternative way of judging the relative performance of the different risk models, I introduce a ranking system which works as follows: For every threshold level (5% or 1%) and every forecasting horizon (1 day or 5 days) the models are ranked according to their performance from the smallest number of rejections to the largest number of rejections. Thus, for example the FP-AST models has ranks 3,1,1 and 1 in the *5% 1d Total*, the *1% 1d Total*, the *5% 5d Total* and the *1% 5d Total* rankings as determined by the respective columns in table 2.4. In a second step, the mean rank achieved by all models is calculated which is displayed in figure 2.2 for the unconditional coverage test. The best performing model, with the lowest rank, is indicated by a white bar, while the worst performing model is indicated by a dark grey bar. As is seen in the upper graph in figure 2.2, the highest rank is achieved by the FP-AST model with an average rank of 1.75, followed by the REVT-ROB model with an average rank of 2. The worst performing model is the least sophisticated FP-QML model. The results when ranking only the performance with respect to the under-estimation of risk for the unconditional coverage test is shown in the bottom graph in figure 2.2. The best performing models are

FP-ST, FP-AST and REVT-ROB all having an average rank of 1.5. Again, the weakest performing model is the FP-QML model. Thus, with respect to the unconditional coverage test, the recently introduced models by Zhu and Galbraith (2010) (FP-AST) as well as Mancini and Trojani (2011) (REVT-ROB) have shown the strongest performance.

Figure 2.2: Ranking Unconditional Coverage Test

The upper graph shows the results of applying a ranking procedure to the results summarized in table 2.4 in the columns labeled *Total*. The procedure works as follows: For every threshold level (5% and 1%) and every forecasting horizon (1 day and 5 days) the models are ranked according to their performance from the smallest number of rejections to the largest number of rejections. In a second step, the mean rank achieved by all models is calculated which is displayed in the upper graph. The best performing model, with the lowest rank, is indicated by a white bar, while the weakest performing model is indicated by a dark grey bar. The bottom graph follows the same procedure, but uses the under-estimation columns labeled *Under* in the summary table.

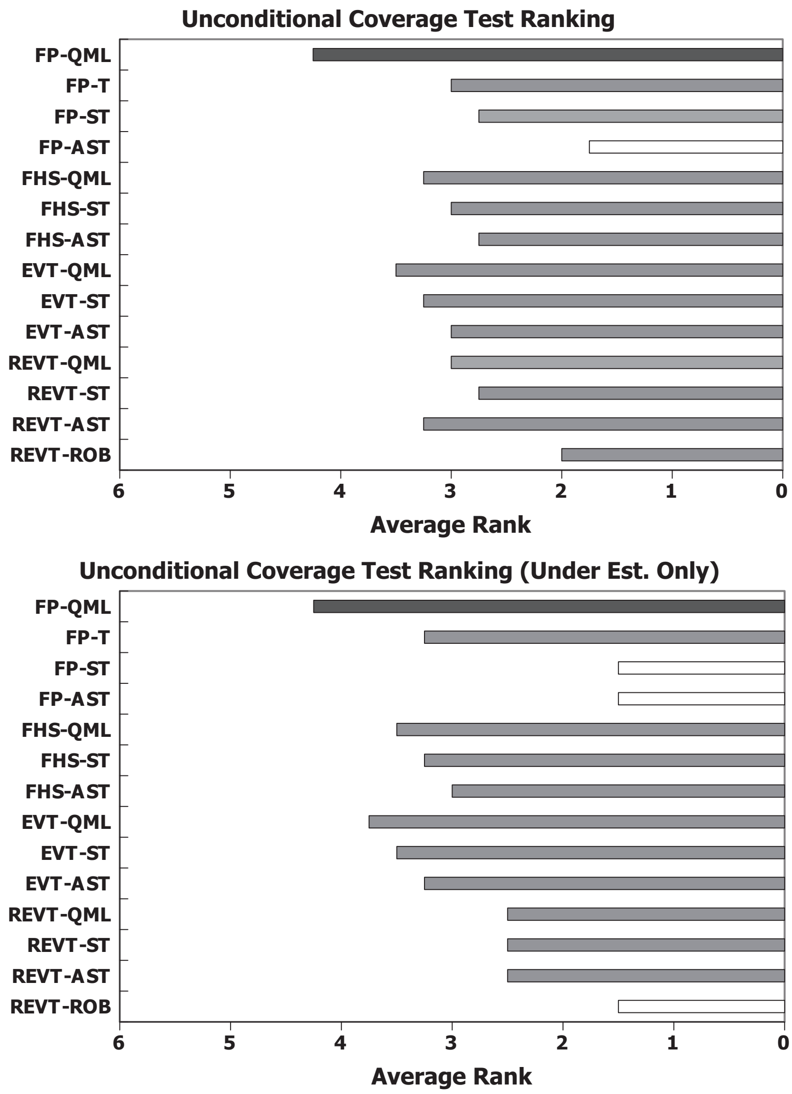


Table 2.2: Unconditional Coverage Test (1-Day Horizon)

This table lists the p-values for the test of unconditional coverage for the 1-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. In order to distinguish under- and over-estimations of risk, the p-values for which the forecasts are too conservative (over-estimation) carry a negative sign. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing models, showing the fewest rejections, are highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0*	0.06	−0.26	0.04*	−0.94	−0.18	0*	0.89	0.68	−0.49	0.31	−0.49	−0.16	3
	1%	0*	0*	0.02*	0*	0*	0*	0*	0.12	0.04*	0.04*	0*	0*	0*	12
FP-T	5%	0*	0*	0.79	0*	0.45	0.89	0*	0.45	0.16	0.45	0.01*	0.68	−0.94	5
	1%	0.01*	0.01*	−0.83	0.09	0.50	0.33	0*	0.81	0.70	−0.40	0*	0.16	−0.95	4
FP-ST	5%	0*	0.34	0*	0.31	0.73	−0.63	0*	0.50	0.22	−1.00	0.59	−0.49	−0.21	3
	1%	0.12	0.59	0.41	−0.95	0.59	0.59	0.03*	0.93	0.70	−0.32	0.05	−0.71	−0.95	1
FP-AST	5%	0*	0.34	0*	0.20	−0.89	−0.73	0*	0.63	0.22	−0.73	0.41	−0.53	−0.40	3
	1%	0.93	0.07	0.04*	−0.95	−0.71	−0.95	0.12	0.93	0.59	−0.14	0.41	−0.32	−0.40	1
FHS-QML	5%	0.03*	0.12	0*	0.09	0.79	0.95	0.03*	0.38	0.28	−0.49	0.89	−0.78	0.25	3
	1%	0.01*	0.04*	0*	0.01*	0.07	0.03*	0*	0.93	0.21	−0.60	0.12	0.16	0*	7
FHS-ST	5%	0.01*	0.10	0*	0.10	0.89	−1.00	0.01*	0.38	0.20	−0.40	−0.89	−0.78	0.31	3
	1%	0.01*	0.04*	0*	0.05	0.33	0.01*	0*	0.81	0.21	−0.50	0.16	0.12	0.01*	6
FHS-AST	5%	0.04*	0.20	0*	0.08	−1.00	−0.83	0.01*	0.34	0.22	−0.49	−0.89	−0.73	0.38	3
	1%	0.03*	0.03*	0*	0.21	0.33	0.02*	0*	0.93	0.21	−0.60	0.27	0.12	0.03*	6
EVT-QML	5%	0.01*	0.03*	0*	0.03*	0.54	0.59	0*	0.31	0.08	−1.00	0.84	−0.58	0.41	5
	1%	0.12	0.04*	0*	0.01*	0.16	0.01*	0.12	−0.95	0.59	−0.32	0.02*	0.16	0*	6
EVT-ST	5%	0*	0.03*	0*	0.02*	0.73	0.68	0*	0.22	0.07	0.89	0.73	−0.53	0.68	5
	1%	0.09	0.12	0*	0*	0.50	0*	0.03*	0.70	0.50	−0.32	0.05	0.16	0*	5
EVT-AST	5%	0.02*	0.06	0*	0.03*	0.89	0.79	0*	0.18	0.07	0.84	0.73	−0.36	0.63	4
	1%	0.09	0.07	0*	0.07	0.50	0*	0.04*	0.59	0.50	−0.32	0.16	0.41	0*	4
REVT-QML	5%	0.04*	0.54	0*	0.01*	0.63	0.63	0.14	0.50	0.59	−0.11	−0.78	−0.68	0.25	3
	1%	0.33	−0.14	0*	0*	−0.60	0*	−0*	−0.05*	−0.14	−0.02*	0.50	0.03*	0*	8
REVT-ST	5%	0.08	0.45	0*	0.01*	0.89	0.59	0.10	0.38	0.59	−0.29	−0.94	−0.78	0.34	2
	1%	0.50	−0.25	0*	0*	−0.40	0*	−0*	−0.14	−0.14	−0.01*	0.93	0*	0*	7
REVT-AST	5%	0.14	0.63	0*	0.02*	0.79	−0.68	0.14	0.34	0.54	−0.08	−1.00	−0.78	0.73	2
	1%	0.12	−0.14	0*	0.01*	0.41	0.05	−0*	−0.05*	−0.03*	−0*	−0.71	0*	0.02*	8
REVT-ROB	5%	0.50	0.73	−0.29	−0.94	0.95	0.84	0.34	−1.00	0.54	−0.68	0.84	0.59	−0.68	0
	1%	0.09	0.81	−0*	0*	0.05	0.12	−0*	−0.10	−0*	−0*	0.07	0*	−0.40	6
NR		16	9	23	16	1	9	22	2	3	5	4	5	10	

Table 2.3: Unconditional Coverage Test (5-Day Horizon)

This table lists the p-values for the test of unconditional coverage for the 5-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. In order to distinguish under- and over-estimations of risk, the p-values for which the forecasts are too conservative (over-estimation) carry a negative sign. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing models, showing the fewest rejections, are highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0.02*	0.32	0.93	0.58	−0.60	−0.71	0.25	0.49	0.49	−0.95	0.49	0.69	−0.31	1
	1%	0.03*	0.49	−0.30	0.01*	0.02*	0.32	0.32	0.02*	0.71	0.03*	0*	0.49	−0.30	6
FP-T	5%	0*	0.15	−0.95	0.15	−0.40	0.81	0.05*	0.58	0.40	0.69	0.32	0.15	−0.49	2
	1%	0.11	0.71	−0.06	0.19	0.49	0.71	0.19	0.11	0.97	0.06	0.02*	0.97	−0.06	1
FP-ST	5%	−0.83	−0.03*	−0*	−0.18	−0.01*	−0.07	−0.13	−0.02*	−0*	−0*	−0.24	−0.02*	−0*	8
	1%	0.97	0.97	−0.02*	0.97	0.97	0.97	0.97	−0.51	−0.51	−0.30	0.49	−0.76	−0.02*	2
FP-AST	5%	0.20	0.58	0.40	0.81	−0.07	−0.60	0.25	0.58	0.32	−0.83	−0.83	−0.60	−0.83	0
	1%	−0.76	0.71	−0.30	0.97	0.97	−0.76	0.32	0.06	0.71	0.71	0.32	−0.76	−0.02*	1
FHS-QML	5%	0.15	0.49	0.03*	0.25	−0.40	−0.83	0.40	0.58	0.25	−0.71	0.69	0.58	0.93	1
	1%	0.71	0.49	0.71	0.03*	0.06	0.06	0.49	0.03*	0.71	0.19	0.11	0.71	−0.76	2
FHS-ST	5%	0.12	0.32	0.06	0.49	−0.49	0.93	0.32	0.58	0.15	−0.95	0.49	0.49	0.58	0
	1%	0.19	0.32	−0.76	0.06	0.11	0.03*	0.19	0.03*	0.71	0.32	0.02*	0.97	−0.76	3
FHS-AST	5%	0.20	0.49	0.03*	0.32	−0.40	0.93	0.40	0.32	0.20	0.81	0.69	0.32	0.58	1
	1%	0.19	0.71	−0.51	0.06	0.11	0.01*	0.32	0.06	0.71	0.49	0.06	0.71	−0.30	1
EVT-QML	5%	0.20	0.32	0.03*	0.25	−0.60	0.81	0.49	0.32	0.25	−0.95	0.81	0.58	0.69	1
	1%	0.32	0.49	0.49	0.03*	0.11	0.03*	0.19	0.06	0.71	0.32	0.06	0.71	−0.76	2
EVT-ST	5%	0.09	0.32	0.03*	0.32	−0.60	0.93	0.58	0.32	0.32	−0.95	0.58	0.40	0.49	1
	1%	0.11	0.49	0.97	0.06	0.11	0.02*	0.32	0.06	0.71	0.32	0.03*	0.97	−0.76	2
EVT-AST	5%	0.12	0.58	0.02*	0.25	−0.40	0.93	0.25	0.40	0.25	0.49	0.81	0.32	0.49	1
	1%	0.11	0.49	0.97	0.02*	0.06	0.02*	0.32	0.03*	0.71	0.49	0.06	0.71	−0.30	3
REVT-QML	5%	0.20	0.40	0.02*	0.25	−0.71	0.69	0.58	0.40	0.25	−0.83	0.81	0.49	0.58	1
	1%	0.32	0.49	0.32	0.06	0.19	0.06	0.71	0.11	0.97	0.49	0.11	0.97	−0.76	0
REVT-ST	5%	0.12	0.32	0.03*	0.32	−0.49	0.81	0.58	0.32	0.25	0.81	0.69	0.40	0.40	1
	1%	0.11	0.49	0.71	0.06	0.06	0.03*	0.71	0.11	0.71	0.71	0.11	0.97	−0.76	1
REVT-AST	5%	0.15	0.69	0.02*	0.25	−0.49	0.93	0.40	0.58	0.25	0.69	0.69	0.32	0.49	1
	1%	0.11	0.49	0.97	0.02*	0.02*	0.11	0.32	0.06	0.71	0.71	0.11	0.71	−0.76	2
REVT-ROB	5%	0.49	−0.71	−0.49	−0.83	−0.83	0.69	−0.95	0.93	0.69	0.81	0.25	0.25	−0.83	0
	1%	0.71	0.97	0.97	0.32	0.19	0.19	0.49	0.19	0.49	−0.51	0.01*	0.97	−0.15	1
NR		3	1	10	5	3	6	1	5	1	2	5	1	3	

Table 2.4: Summary Unconditional Coverage Test

This table summarizes the results for the unconditional coverage test by listing and aggregating the number of rejections (NR) at the 5% level of significance for each threshold level (5% and 1%) as well as for both considered forecasting horizons (1d and 5d). Thereby, the number of rejections due to an under- respectively an over-estimation of risk are separately shown for each threshold level and forecasting horizon combination. The best performing model, showing the fewest rejections overall (column: *All Total*), is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

	5% 1d Under	5% 1d Over	5% 1d Total	1% 1d Under	1% 1d Over	1% 1d Total	All 1d Total	5% 5d Under	5% 5d Over	5% 5d Total	1% 5d Under	1% 5d Over	1% 5d Total	All 5d Total	All Total	All Under	All Over
FP-QML	3	0	3	12	0	12	15	1	0	1	6	0	6	7	22	22	0
FP-T	5	0	5	4	0	4	9	2	0	2	1	0	1	3	12	12	0
FP-ST	3	0	3	1	0	1	4	0	8	8	0	2	2	10	14	4	10
FP-AST	3	0	3	1	0	1	4	0	0	0	0	1	1	1	5	4	1
FHS-QML	3	0	3	7	0	7	10	1	0	1	2	0	2	3	13	13	0
FHS-ST	3	0	3	6	0	6	9	0	0	0	3	0	3	3	12	12	0
FHS-AST	3	0	3	6	0	6	9	1	0	1	1	0	1	2	11	11	0
EVT-QML	5	0	5	6	0	6	11	1	0	1	2	0	2	3	14	14	0
EVT-ST	5	0	5	5	0	5	10	1	0	1	2	0	2	3	13	13	0
EVT-AST	4	0	4	4	0	4	8	1	0	1	3	0	3	4	12	12	0
REVT-QML	3	0	3	5	3	8	11	1	0	1	0	0	0	1	12	9	3
REVT-ST	2	0	2	5	2	7	9	1	0	1	1	0	1	2	11	9	2
REVT-AST	2	0	2	4	4	8	10	1	0	1	2	0	2	3	13	9	4
REVT-ROB	0	0	0	2	4	6	6	0	0	0	1	0	1	1	7	3	4

The next test under consideration is the test of independence, which analyzes whether the VaR violations are independent over time. The results for the 1-day horizon are listed in table 2.5. Again, p-values with the suffix * indicate cases for which the null hypothesis of independent VaR violations can be rejected at the 5% level of significance. The results are summarized in the same way as for the unconditional coverage test by adding up the number of rejections at the 5% level of significance and are shown in table 2.7. The best models at the 1-day horizon are the less sophisticated FP-QML and FP-T models with only 2 rejections (column: *1d Total*). However, many models perform similarly well with all other models having only 3 or 4 rejections except the REVT-AST and REVT-ROB models with 5 respectively 8 rejections. When analyzing the rejections with respect to the indices shown in the bottom row of table 2.5, it is again interesting to see that certain indices seem to be more challenging than others. In particular the CRED and CUVA indices with 25 respectively 12 rejections are causing difficulties with respect to the test of independence.

Moving to the 5-day horizon, the results are similar and all models perform well with the results ranging from 0 to 2 rejections (column: *5d Total*). The best performing models are FP-AST and REVT-ROB with 0 rejections. The overall results show that the FP-T and FP-AST models perform the best with only 3 rejections each (column: *All Total*). The worst performing models are FHS-AST, EVT-AST, REVT-QML with 6 rejections each as well as the REVT-ROB model with 8 rejections.

Table 2.5: Independence Test (1-Day Horizon)

This table lists the p-values for the test of independence for the 1-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing models, showing the fewest rejections, are highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0.18	0.18	0.20	0.71	1.00	0.07	0.71	0.83	0.04*	0.71	0.13	0.55	0.12	1
	1%	0.81	0.56	0.19	0.41	0.74	0.01*	0.95	0.71	0.24	0.21	0.25	0.06	0.65	1
FP-T	5%	0.09	0.17	0.42	0.28	0.60	0.01*	0.58	0.60	0.12	0.71	0.14	0.49	0.08	1
	1%	0.17	0.31	0.36	0.23	0.57	0*	0.10	0.33	0.53	0.41	0.97	0.25	0.08	1
FP-ST	5%	0.32	0.20	0.44	0.34	0.31	0.01*	0.98	0.58	0.10	0.72	0.23	0.36	0.02*	2
	1%	0.24	0.31	0.29	0.35	0.55	0*	0.26	0.34	0.53	0.42	0.79	0.37	0.46	1
FP-AST	5%	0.37	0.20	0.37	0.58	0.52	0.01*	0.97	0.72	0.10	0.84	0.29	0.24	0.03*	2
	1%	0.34	0.76	0.21	0.35	0.37	0*	0.19	0.34	0.55	0.46	0.59	0.42	0.36	1
FHS-QML	5%	0.18	0.04*	0.18	0.94	0.88	0.02*	0.92	0.46	0.08	0.71	0.41	0.68	0.05*	3
	1%	0.91	0.81	0.61	0.18	0.76	0*	0.97	0.34	0.67	0.38	0.71	0.25	0.96	1
FHS-ST	5%	0.13	0.03*	0.50	0.35	0.41	0.01*	0.44	0.46	0.10	0.98	0.34	0.47	0.02*	3
	1%	0.19	0.81	0.71	0.22	0.28	0*	0.44	0.33	0.67	0.39	0.69	0.24	0.89	1
FHS-AST	5%	0.20	0.03*	0.44	0.39	0.56	0.02*	0.44	0.67	0.10	0.77	0.21	0.45	0.04*	3
	1%	0.20	0.84	0.65	0.26	0.28	0*	0.38	0.34	0.67	0.38	0.64	0.24	0.84	1
EVT-QML	5%	0.33	0.06	0.16	0.90	0.77	0.04*	0.85	0.69	0.10	0.72	0.17	0.59	0.03*	2
	1%	0.24	0.81	0.87	0.19	0.69	0*	0.24	0.35	0.55	0.42	0.86	0.25	0.90	1
EVT-ST	5%	0.30	0.03*	0.46	0.40	0.47	0.01*	0.66	0.56	0.11	0.92	0.20	0.57	0.01*	3
	1%	0.23	0.71	0.97	0.16	0.57	0*	0.26	0.32	0.57	0.42	0.79	0.25	0.40	1
EVT-AST	5%	0.25	0.02*	0.34	0.35	0.60	0.03*	0.66	0.61	0.11	0.85	0.12	0.31	0.01*	3
	1%	0.23	0.76	0.94	0.22	0.30	0*	0.24	0.31	0.57	0.42	0.69	0.29	0.38	1
REVT-QML	5%	0.06	0.05*	0.12	0.93	0.80	0.02*	0.65	0.79	0.08	0.98	0.31	0.64	0.03*	3
	1%	0.28	0.46	0.09	0.10	0.40	0*	0.57	0.50	0.46	0.52	0.30	0.20	0.70	1
REVT-ST	5%	0.14	0.06	0.94	0.48	0.60	0.01*	0.35	0.65	0.08	0.91	0.54	0.47	0.02*	2
	1%	0.30	0.43	0.16	0.12	0.36	0*	0.57	0.46	0.28	0.54	0.34	0.16	0.54	1
REVT-AST	5%	0.18	0.04*	0.37	0.27	0.45	0.03*	0.32	0.32	0.09	0.44	0.38	0.19	0.01*	3
	1%	0.24	0.46	0.15	0.18	0.02*	0*	0.57	0.50	0.51	0.57	0.37	0.96	0.06	2
REVT-ROB	5%	0.73	0*	0*	0*	0.15	0.28	0.07	0.02*	0.05*	0.64	0.43	0.98	0.87	5
	1%	0.74	0.50	0.70	0.01*	0.05*	0.19	0.59	0.47	0.58	0.62	0.21	0.52	0*	3
NR		0	8	1	2	2	25	0	1	2	0	0	0	12	0

Table 2.6: Independence Test (5-Day Horizon)

This table lists the p-values for the test of independence for the 5-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing models, showing the fewest rejections, are highlighted in bold and the weakest performing models, showing the highest number of rejections, are highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0.51	0.97	0.63	0.24	0.05*	0.18	0.50	0.85	0.21	0.58	0.43	0.74	0.77	1
	1%	0.44	0.60	0.77	0.39	0.42	0.01*	0.57	0.42	0.63	0.29	0.31	0.60	0.77	1
FP-T	5%	0.75	0.68	0.24	0.42	0.06	0.31	0.65	0.68	0.19	0.35	0.50	0.86	0.39	0
	1%	0.50	0.63	0.85	0.53	0.60	0*	0.53	0.50	0.67	0.25	0.42	0.67	0.85	1
FP-ST	5%	0.38	0.13	0.62	0.08	0.18	0.04*	0.61	0.15	0.31	0.19	0.71	0.99	0.57	1
	1%	0.67	0.67	0.89	0.67	0.67	0*	0.67	0.74	0.74	0.77	0.60	0.70	0.89	1
FP-AST	5%	0.92	0.80	0.21	0.29	0.11	0.16	0.98	0.80	0.17	0.53	0.53	0.43	0.53	0
	1%	0.70	0.63	0.77	0.67	0.67	0.06	0.57	0.47	0.63	0.63	0.57	0.70	0.89	0
FHS-QML	5%	0.86	0.85	0.57	0.15	0.06	0.21	0.48	0.68	0.15	0.18	0.35	0.80	0.63	0
	1%	0.63	0.60	0.63	0.44	0.47	0.02*	0.60	0.44	0.63	0.18	0.50	0.63	0.06	1
FHS-ST	5%	0.81	0.97	0.23	0.21	0.05	0.27	0.97	0.68	0.12	0.24	0.41	0.85	0.80	0
	1%	0.53	0.57	0.70	0.47	0.50	0*	0.53	0.44	0.63	0.15	0.42	0.67	0.06	1
FHS-AST	5%	0.92	0.85	0.29	0.17	0.06	0.27	0.48	0.97	0.14	0.31	0.71	0.97	0.39	0
	1%	0.53	0.63	0.74	0.47	0.50	0*	0.57	0.47	0.63	0.12	0.47	0.63	0.03*	2
EVT-QML	5%	0.92	0.97	0.57	0.15	0.05*	0.31	0.64	0.54	0.15	0.24	0.31	0.80	0.74	1
	1%	0.57	0.60	0.12	0.44	0.50	0.03*	0.53	0.47	0.63	0.15	0.47	0.63	0.06	1
EVT-ST	5%	0.76	0.97	0.29	0.17	0.46	0.10	0.80	0.97	0.17	0.24	0.37	0.91	0.43	0
	1%	0.50	0.60	0.67	0.47	0.50	0*	0.57	0.47	0.63	0.15	0.44	0.67	0.06	1
EVT-AST	5%	0.81	0.80	0.33	0.15	0.06	0.27	0.58	0.91	0.15	0.43	0.66	0.97	0.43	0
	1%	0.50	0.60	0.08	0.42	0.47	0*	0.57	0.44	0.63	0.12	0.47	0.63	0.03*	2
REVT-QML	5%	0.92	0.91	0.68	0.15	0.04*	0.35	0.68	0.59	0.15	0.21	0.31	0.85	0.39	1
	1%	0.57	0.60	0.15	0.47	0.53	0.02*	0.63	0.50	0.67	0.12	0.50	0.67	0.06	1
REVT-ST	5%	0.81	0.97	0.29	0.17	0.05	0.11	0.80	0.97	0.15	0.31	0.35	0.91	0.48	0
	1%	0.50	0.60	0.63	0.47	0.47	0*	0.63	0.50	0.63	0.10	0.50	0.67	0.06	1
REVT-AST	5%	0.86	0.74	0.33	0.15	0.05	0.27	0.48	0.68	0.15	0.35	0.71	0.97	0.43	0
	1%	0.50	0.60	0.08	0.42	0.42	0*	0.57	0.47	0.63	0.10	0.50	0.63	0.06	1
REVT-ROB	5%	0.64	0.48	0.50	0.95	0.38	0.26	0.58	0.32	0.26	0.69	1.00	0.98	0.95	0
	1%	0.63	0.67	0.67	0.57	0.53	0.53	0.60	0.53	0.60	0.74	0.39	0.67	0.81	0
NR		0	0	0	0	3	13	0	0	0	0	0	0	2	

Table 2.7: Summary Independence Test

This table summarizes the results for the test of independence by listing and aggregating the number of rejections (NR) at the 5% level of significance for each threshold level (5% and 1%) as well as for both considered forecasting horizons (1d and 5d). The best performing models, showing the fewest rejections overall (column: *All Total*), are highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

	5% 1d	1% 1d	1d Total	5% 5d	1% 5d	5d Total	All Total
FP-QML	1	1	2	1	1	2	4
FP-T	1	1	2	0	1	1	3
FP-ST	2	1	3	1	1	2	5
FP-AST	2	1	3	0	0	0	3
FHS-QML	3	1	4	0	1	1	5
FHS-ST	3	1	4	0	1	1	5
FHS-AST	3	1	4	0	2	2	6
EVT-QML	2	1	3	1	1	2	5
EVT-ST	3	1	4	0	1	1	5
EVT-AST	3	1	4	0	2	2	6
REVT-QML	3	1	4	1	1	2	6
REVT-ST	2	1	3	0	1	1	4
REVT-AST	3	2	5	0	1	1	6
REVT-ROB	5	3	8	0	0	0	8

The last test related to the prediction of VaR is the test of conditional coverage, which tests jointly for the correct number of VaR predictions as well as the independence of the observations. The results for the 1-day horizon are shown in table 2.8 and p-values with the suffix * highlight cases for which the null hypothesis can be rejected at the 5% level of significance. The summary of the results is found in table 2.10.

The findings for the 1-day horizon of the test of conditional coverage to some extent reflect those of the test of unconditional coverage for the 1-day horizon which can be expected by the definition of the test. The best model with only 5 rejections is the FP-AST model, followed by the FP-ST and FHS-AST models with 7 rejections each at the 5% level of significance (column: *1d Total*). The weakest models are the FP-QML and the EVT-ST models, with 12 rejections each. Again, the robust models perform relatively well for the 5% threshold and less so for the 1% threshold.

Moving to the 5-day horizon for which the results are shown in table 2.9, it can be seen that all models except the FP-ST perform relatively well, with the best model FP-AST showing 0 rejections and the FHS-QML, FHS-AST, EVT-QML, EVT-ST, REVT-ST as well as the REVT-ROB models only showing 1 rejection each (column: *5d Total*). The outlier here is clearly the FP-ST model with 9 rejections. Analyzing the overall results which are found by adding the rejections for the 1-day and the 5-day horizon forecasts, the FP-AST model with 5 rejections followed by the FHS-AST model with 8 rejections and the EVT-QML as well as the FHS-QML models with 9 respectively 10 rejections each are the best performing models. The weakest model overall is the FP-QML model with 18 rejections and the second weakest model is the FP-ST model with 16 rejections. However, as analyzed in the section about the test of unconditional coverage, the weak performance of the FP-ST model is mainly driven by a strong over-estimation of risk at the 5% threshold for the 5-day forecasting horizon. Using the same ranking procedure as for the test of unconditional coverage, the relative performance of the different models have been ranked and the results are shown in figure 2.3. For the conditional coverage test ranking, the results from the summary table are confirmed. The FP-AST model has the lowest rank (1.5), followed by the FHS-AST model (2.25) and the EVT-QML (2.5).

Table 2.8: Conditional Coverage Test (1-Day Horizon)

This table lists the p-values for the conditional coverage test for the 1-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing model, showing the fewest rejections, is highlighted in bold and the weakest performing models, showing the highest number of rejections, are highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0*	0.07	0.23	0.12	1.00	0.08	0.01*	0.97	0.11	0.73	0.19	0.65	0.11	2
	1%	0*	0*	0.03*	0*	0*	0*	0*	0.29	0.06	0.05	0*	0*	0*	10
FP-T	5%	0*	0.01*	0.70	0.01*	0.66	0.05*	0*	0.66	0.11	0.70	0.01*	0.73	0.21	6
	1%	0.01*	0.02*	0.64	0.12	0.68	0*	0*	0.60	0.76	0.50	0*	0.19	0.21	5
FP-ST	5%	0*	0.28	0.01*	0.38	0.57	0.04*	0*	0.68	0.12	0.94	0.42	0.52	0.03*	5
	1%	0.15	0.51	0.40	0.64	0.72	0*	0.04*	0.63	0.76	0.44	0.14	0.63	0.76	2
FP-AST	5%	0*	0.28	0*	0.38	0.80	0.05*	0*	0.84	0.12	0.92	0.41	0.41	0.07	4
	1%	0.63	0.18	0.05	0.64	0.63	0*	0.13	0.63	0.72	0.26	0.62	0.44	0.46	1
FHS-QML	5%	0.04*	0.04*	0*	0.24	0.95	0.08	0.08	0.51	0.12	0.73	0.71	0.88	0.07	3
	1%	0.03*	0.11	0*	0.01*	0.18	0*	0*	0.63	0.41	0.60	0.29	0.19	0.01*	6
FHS-ST	5%	0.01*	0.02*	0*	0.17	0.71	0.04*	0.03*	0.51	0.12	0.70	0.63	0.74	0.04*	6
	1%	0.02*	0.11	0*	0.07	0.35	0*	0*	0.60	0.41	0.55	0.35	0.15	0.04*	5
FHS-AST	5%	0.06	0.04*	0*	0.15	0.84	0.06	0.03*	0.58	0.12	0.75	0.46	0.71	0.07	3
	1%	0.04*	0.08	0*	0.24	0.35	0*	0*	0.63	0.41	0.60	0.48	0.15	0.08	4
EVT-QML	5%	0.02*	0.01*	0*	0.10	0.79	0.11	0.02*	0.55	0.05	0.94	0.39	0.74	0.07	4
	1%	0.15	0.11	0*	0.02*	0.35	0*	0.15	0.64	0.72	0.44	0.06	0.19	0*	4
EVT-ST	5%	0.01*	0.01*	0*	0.04*	0.73	0.03*	0.01*	0.40	0.05	0.99	0.41	0.70	0.03*	7
	1%	0.12	0.29	0*	0*	0.68	0*	0.04*	0.56	0.68	0.44	0.14	0.19	0*	5
EVT-AST	5%	0.03*	0.01*	0*	0.06	0.87	0.10	0.01*	0.35	0.05	0.96	0.27	0.40	0.03*	5
	1%	0.12	0.18	0*	0.09	0.46	0*	0.06	0.51	0.68	0.44	0.35	0.40	0*	3
REVT-QML	5%	0.02*	0.12	0*	0.03*	0.87	0.06	0.30	0.77	0.19	0.27	0.57	0.82	0.04*	4
	1%	0.35	0.26	0*	0*	0.61	0*	0.02*	0.11	0.26	0.06	0.46	0.04*	0*	6
REVT-ST	5%	0.07	0.12	0*	0.03*	0.87	0.03*	0.17	0.61	0.19	0.57	0.82	0.74	0.04*	4
	1%	0.46	0.38	0.01*	0*	0.46	0*	0.02*	0.26	0.19	0.04*	0.63	0.01*	0*	7
REVT-AST	5%	0.13	0.11	0*	0.03*	0.73	0.08	0.20	0.39	0.19	0.16	0.68	0.41	0.03*	3
	1%	0.15	0.26	0*	0.01*	0.05*	0*	0.02*	0.11	0.08	0.02*	0.63	0.01*	0.01*	8
REVT-ROB	5%	0.75	0.01*	0*	0*	0.36	0.55	0.13	0.07	0.12	0.82	0.72	0.86	0.90	3
	1%	0.23	0.78	0*	0*	0.02*	0.13	0.01*	0.20	0.01*	0*	0.09	0*	0.01*	8
NR		15	10	23	14	3	19	20	0	1	3	3	5	17	

Table 2.9: Conditional Coverage Test (5-Day Horizon)

This table lists the p-values for the conditional coverage test for the 5-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing model, showing the fewest rejections, is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0.04*	0.61	0.89	0.43	0.12	0.38	0.42	0.77	0.36	0.86	0.58	0.87	0.58	1
	1%	0.07	0.69	0.56	0.02*	0.04*	0.02*	0.52	0.04*	0.83	0.06	0*	0.69	0.56	5
FP-T	5%	0.02*	0.33	0.50	0.26	0.12	0.58	0.12	0.79	0.30	0.60	0.49	0.35	0.55	1
	1%	0.22	0.83	0.18	0.35	0.69	0.01*	0.35	0.22	0.91	0.09	0.04*	0.91	0.18	2
FP-ST	5%	0.67	0.03*	0*	0.09	0.01*	0.02*	0.28	0.02*	0*	0*	0.47	0.06	0*	8
	1%	0.91	0.91	0.07	0.91	0.91	0.01*	0.91	0.76	0.76	0.56	0.69	0.89	0.07	1
FP-AST	5%	0.43	0.83	0.32	0.55	0.05	0.32	0.52	0.83	0.24	0.80	0.80	0.64	0.80	0
	1%	0.89	0.83	0.56	0.91	0.91	0.16	0.52	0.13	0.83	0.83	0.52	0.89	0.07	0
FHS-QML	5%	0.35	0.77	0.09	0.19	0.12	0.45	0.54	0.79	0.19	0.38	0.60	0.83	0.89	0
	1%	0.83	0.69	0.83	0.07	0.13	0.01*	0.69	0.07	0.83	0.17	0.22	0.83	0.16	1
FHS-ST	5%	0.28	0.61	0.08	0.36	0.12	0.55	0.61	0.79	0.11	0.50	0.56	0.77	0.83	0
	1%	0.35	0.52	0.89	0.13	0.22	0*	0.35	0.07	0.83	0.21	0.04*	0.91	0.16	2
FHS-AST	5%	0.43	0.77	0.06	0.24	0.12	0.55	0.54	0.61	0.14	0.58	0.86	0.61	0.59	0
	1%	0.35	0.83	0.76	0.13	0.22	0*	0.52	0.13	0.83	0.24	0.13	0.83	0.05	1
EVT-QML	5%	0.43	0.61	0.09	0.19	0.12	0.58	0.70	0.51	0.19	0.50	0.58	0.83	0.87	0
	1%	0.52	0.69	0.24	0.07	0.22	0.01*	0.35	0.13	0.83	0.21	0.13	0.83	0.16	1
EVT-ST	5%	0.22	0.61	0.06	0.24	0.66	0.25	0.83	0.61	0.24	0.50	0.57	0.69	0.58	0
	1%	0.22	0.69	0.91	0.13	0.22	0*	0.52	0.13	0.83	0.21	0.07	0.91	0.16	1
EVT-AST	5%	0.28	0.83	0.05*	0.19	0.12	0.55	0.45	0.69	0.19	0.58	0.88	0.61	0.58	1
	1%	0.22	0.69	0.21	0.04*	0.13	0*	0.52	0.07	0.83	0.24	0.13	0.83	0.05	2
REVT-QML	5%	0.43	0.69	0.05*	0.19	0.12	0.60	0.79	0.60	0.19	0.45	0.58	0.77	0.59	1
	1%	0.52	0.69	0.21	0.13	0.35	0.01*	0.83	0.22	0.91	0.24	0.22	0.91	0.16	1
REVT-ST	5%	0.28	0.61	0.06	0.24	0.12	0.28	0.83	0.61	0.19	0.58	0.60	0.69	0.54	0
	1%	0.22	0.69	0.83	0.13	0.13	0*	0.83	0.22	0.83	0.24	0.22	0.91	0.16	1
REVT-AST	5%	0.35	0.87	0.05*	0.19	0.12	0.55	0.54	0.79	0.19	0.60	0.86	0.61	0.58	1
	1%	0.22	0.69	0.21	0.04*	0.04*	0*	0.52	0.13	0.83	0.24	0.22	0.83	0.16	3
REVT-ROB	5%	0.70	0.73	0.63	0.97	0.67	0.49	0.86	0.61	0.49	0.89	0.52	0.52	0.97	0
	1%	0.83	0.91	0.91	0.52	0.35	0.35	0.69	0.35	0.69	0.76	0.02*	0.91	0.35	1
NR		2	1	4	3	3	13	0	2	1	1	4	0	1	

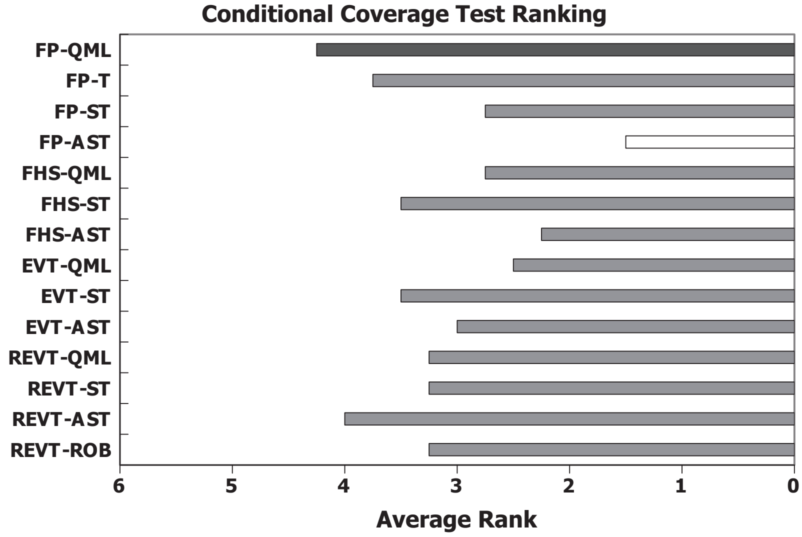
Table 2.10: Summary Conditional Coverage Test

This table summarizes the results for the conditional coverage test by listing and aggregating the number of rejections (NR) at the 5% level of significance for each threshold level (5% and 1%) as well as for both considered forecasting horizons (1d and 5d). The best performing model, showing the fewest rejections overall (column *All Total*), is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

	5% 1d	1% 1d	1d Total	5% 5d	1% 5d	5d Total	All Total
FP-QML	2	10	12	1	5	6	18
FP-T	6	5	11	1	2	3	14
FP-ST	5	2	7	8	1	9	16
FP-AST	4	1	5	0	0	0	5
FHS-QML	3	6	9	0	1	1	10
FHS-ST	6	5	11	0	2	2	13
FHS-AST	3	4	7	0	1	1	8
EVT-QML	4	4	8	0	1	1	9
EVT-ST	7	5	12	0	1	1	13
EVT-AST	5	3	8	1	2	3	11
REVT-QML	4	6	10	1	1	2	12
REVT-ST	4	7	11	0	1	1	12
REVT-AST	3	8	11	1	3	4	15
REVT-ROB	3	8	11	0	1	1	12

Figure 2.3: Ranking Conditional Coverage Test

The graph shows the results of applying a ranking procedure to the results summarized in table 2.10 in the columns labeled *Total*. The procedure works as follows: For every threshold level (5% and 1%) and every forecasting horizon (1 day and 5 days) the models are ranked according to their performance from the smallest number of rejections to the largest number of rejections. In a second step, the mean rank achieved by all models is calculated which is displayed in the graph. The best performing model, with the lowest rank, is indicated by a white bar, while the weakest performing model is indicated by a dark grey bar.



The results of the tests related to the VaR measurement are interesting in different ways. They indicate that the more sophisticated models perform better at predicting VaR for a 1-day horizon and at the 5-day horizon. However, the differences in the performances of the models become smaller at the 5-day horizon. As it will be explained in section 2.5 and shown in the appendix, this is likely the result of lighter tails of the return distributions exhibited at the 5-day horizon compared to the 1-day horizon. Overall, the FP-AST model based on the approach proposed in Zhu and Galbraith (2011), has proven to perform the best on average in the context of forecasting VaR for the various indices. The sophisticated REVT-ROB model was the second best performing model regarding the unconditional coverage test. However, with regards to the test of independence, the model revealed some weaknesses at the 1-day forecasting horizon. With regards to the REVT-ROB model it is noteworthy that with regards to the EQMK index it was the best performing model together with the REVT-ST and REVT-AST models, showing no rejections at all with regards to the VaR related tests.

In order to gain insights into the economic implications of using the different risk models, the following case study analysis is conducted which will be further expanded in subsection 2.4.2. Thereby, the view of a risk manager is taken who intends to decide on which risk model she should choose in order to measure the risk of different kinds of portfolios, represented by the 13 indices used in this study. The focus will be on making 1-day risk forecasts. Thereby, with regards to the VaR risk measure an important economic variable is the number of days which can be expected on average between VaR violations when using different models on the different indices. The expected number of days between VaR violations for 1-day forecasts is given by $1/p$, where p is the threshold level. Thus, using threshold levels of 1% and 5%, the expected average days between VaR violations are 100 respectively 20 days. The graphs in the following figure (2.4) depict box-plot diagrams of the average days between VaR violations for each model when applied to the 13 indices[20]. The upper graph is related to the 1% threshold level, for which violations are expected to occur every 100th day and the bottom graph is related to the 5% threshold level, for which violations are expected to occur every 20th day if the models are correctly predicting VaR[21]. With regards to the 1% threshold (upper graph), it can be seen that for most models when applied on the 13 indices, the median of the average days between VaR violations is too low. Thus, the models are under-estimating VaR for over half of the indices under consideration. The weakest performing model is the FP-QML model with a median value of 56.9 days, far below the expected 100 days. The best performing models are the FP-AST and REVT-ST which show a median value of 98.7 days in the given setting. However, comparing the dispersion of the average days between VaR Violations between these

[20] The bottom line of each box represents the first quartile of the distribution, the line inside the box corresponds to the second quartile (median) and the top line of the box is the third quartile. The extensions (whiskers) at the bottom and the top of the boxes represent the minimum and maximum values of the observed data.

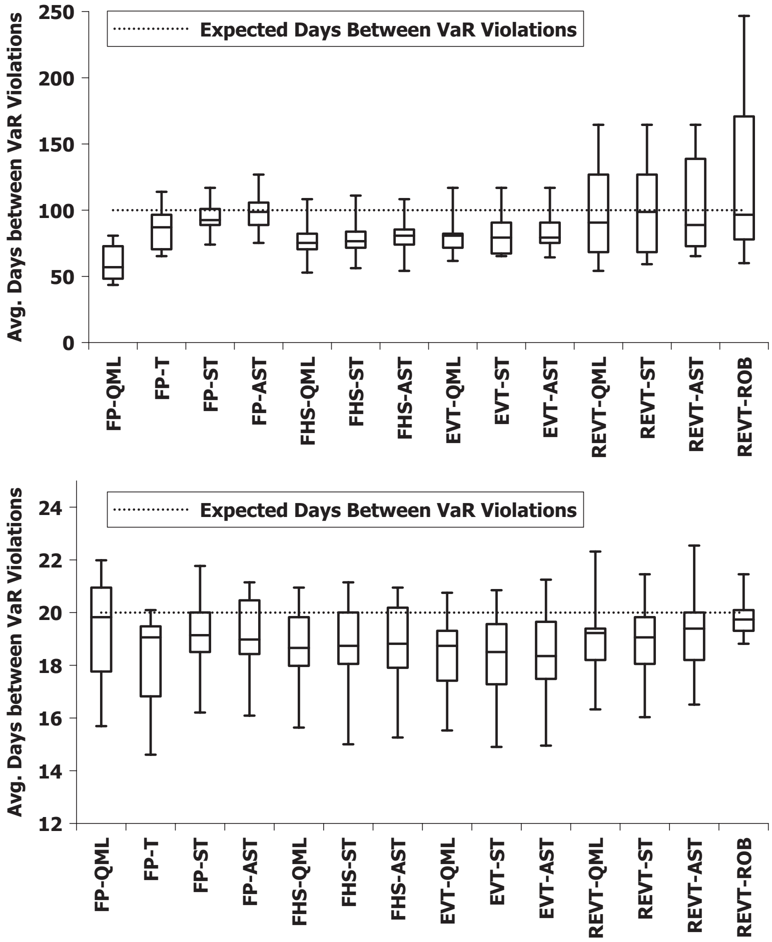
[21] The corresponding statistical test is the test of unconditional coverage discussed at the beginning of this subsection.

two models, the advantage of the FP-AST model is evident as the dispersion is much lower and almost evenly distributed around the 100-day mark. Moving to the 5% threshold in the bottom graph, it is seen that in general the models perform better, with the EVT-AST model showing the lowest median of 18.4 days whereby 20 days are expected. Surprisingly, the best model with regards to the median value is the least sophisticated FP-QML model (19.8 days) closely followed by the REVT-ROB model (19.7 days). However, again the dispersions of the estimates for the two models are very different, making REVT-ROB more attractive[22]. From the risk managers point of view this analysis gives an idea of the economic implications which may occur when choosing a less sophisticated model. The analysis also reveals that the choice of the model is more important when considering the far left tail of the distribution ($p = 1\%$) compared to the 5% threshold. Given that the FP-AST model performs very well under the 1% threshold and is not significantly inferior to most other models at the 5% threshold it may be considered as the best suited model in the given context. In the next section the focus will be on the performance of forecasting the expected shortfall (ES).

[22]From a statistical point of view this is seen in the results of the unconditional coverage ratio test summarized in table 2.4.

Figure 2.4: Case Study - Days Between VaR Violations

The graphs in this figure depict box-plot diagrams of the average days between VaR violations for each model when applied to the 13 indices considered in this study. Thereby, the upper graph is related to the 1% threshold level, for which violations are expected to occur every 100th day and the bottom graph is related to the 5% threshold level, for which violations are expected to occur every 20th day if the models are correctly predicting VaR.



2.4.2 Testing the Performance of Expected Shortfall (ES) Predictions

In this subsection the performance of the various risk models at predicting the expected shortfall (ES) is being analyzed. The test applied is the mean zero test as described in subsection 2.3.2. As in the tables for the VaR related tests, tables 2.11 and 2.12 show the p-values for the null hypothesis that the mean of the exceedance residuals as defined in equation 2.16 is equal to zero, whereby a two-sided test is applied. The suffix * highlights p-values for which the null hypothesis can be rejected at the 5% level of significance. In order to distinguish over- and under-estimations, the p-values related to an over-estimation of risk, carry a negative sign. The results for the 1-day horizon forecasts are listed in table 2.11 and similar to the previous tests, the results are summarized by adding the number of rejections at the 5% significance level whereby the summarized results are shown in table 2.13.

The best performing model for the 1-day horizon is FP-ST with no rejection at all, followed by the EVT-ST and EVT-AST models with 5 rejections each and the FHS-QML model with 6 rejections (column: *All 1d Total*). The weakest model is the FP-QML model with 26 rejections, giving a clear indication that the assumption of normally distributed innovations is not feasible when forecasting the expected shortfall. An interesting result is found for the FP-AST model, where 11 rejections are found. However, only 2 rejections are due to an under-estimation of ES, meaning that if a one sided test would have been performed, as done for example in McNeil and Frey (2000), the model would have shown the second best performance with regards to the ES mean zero test at the 1-day horizon. When looking at the rejections in the ES mean zero test with respect to the indices, EVAL with 18 rejections poses the biggest challenge to the risk models, similar to the results for the unconditional coverage test.

Moving to the 5-day horizon, the results are similar to the findings for the unconditional coverage test where the performances of the various risk models improve and become more homogenous. The best model at forecasting the ES at the 5-day horizon in the given setting is the REVT-ROB model with one rejection, followed by the REVT-QML model with 2 rejections (column: *All 5d Total*). The worst model with 7 rejections is surprisingly the FP-AST model which has been one of the best performing models in the tests so far. However, 5 out of the 7 rejections are the result of a significant over-estimations of the ES. With regards to the indices, CRED was found to have the highest number of rejections at the 5-day horizon, namely 11.

When adding up the rejection numbers for the 1-day and 5-day horizons, the best performing model with 5 rejections is FP-ST (column: *All Total*). The worst performing model is FP-QML with 31 rejections. The same ranking procedure applied to the results of the test of unconditional coverage, is also applied to the ES mean zero test and the results are shown in figure 2.5. Again, the results of two versions of the ranking procedure are shown. In the upper graph in figure 2.5, the ranking is applied to the total results of the ES mean zero test whereas in the bottom graph, the ranking only considers the underestimation of risk[23]. As can be seen, the best ranked models when considering over- and underestimation of ES are FP-ST (rank 2) followed by FHS-QML, EVT-QML, EVT-ST and EVT-AST (rank 2.5). Further, when only considering the under-estimation of ES, the best performing models are FP-ST and FP-AST with ranks of 1.25 and 2 respectively.

The analysis of the ES prediction performance for the different models has shown, that it is crucial whether a one-sided or two-sided test is applied. If only the under-estimation of risk is a concern, the best performing model in the VaR related tests, FP-AST, is the second best model in the ES mean zero test. Therefore, the FP-AST model would likely be considered as the best performing model for the given indices and time frame when forecasting VaR and ES. However, when the over-estimation of risk is also a concern, FP-ST as well as EVT-based methods (EVT-QML, EVT-ST and EVT-AST) show the strongest performance with regards to the ES estimation. An attempt to find the best performing model considering all tests in this study is conducted in the concluding section. Similar to the analysis of the VaR prediction performance, the differences in performance among the models become smaller once a 5-day horizon is considered.

[23] The average rank of a model in the underestimation category is found by taking the average of the *5% 1d Under*, the *1% 1d Under*, the *5% 5d Under* and the *1% 5d Under* rankings as determined by ordering the results in the respective columns in table 2.13 in ascending order.

Table 2.11: Expected Shortfall Mean Zero Test (1-Day Horizon)

This table lists the p-values for the expected shortfall mean zero test for the 1-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. In order to distinguish under- and over-estimations of risk, the p-values for which the forecasts are too conservative (over-estimation) carry a negative sign. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing model, showing the fewest rejections, is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	13
	1%	0*	0*	0*	0*	0*	0*	0.02*	0*	0*	0*	0*	0*	0*	13
FP-T	5%	0.49	0*	−0*	−0.10	−0.13	−0.36	0*	0*	0.01*	−0.36	−0.23	−0.10	−0*	6
	1%	−0.06	0.94	−0.03*	−0*	−0.22	−0.02*	−0.95	0.20	0.36	−0.07	−0*	−0*	−0*	6
FP-ST	5%	−0.93	0.55	−0.09	−0.30	0.94	0.60	0.19	−0.92	1.00	−0.24	0.98	−0.30	−0.51	0
	1%	0.97	0.53	−0.77	−0.42	0.80	0.82	0.98	0.14	0.40	0.82	−0.21	−0.14	−0.07	0
FP-AST	5%	−0*	0.10	−0.90	−0*	−0.04*	−0.06	0.56	0.48	0.93	−0*	−0.02*	−0*	−0.38	6
	1%	−0*	0.08	0.02*	−0.11	−0.22	−0.17	0.92	0.01*	0.26	−0.31	−0*	−0*	0.99	5
FHS-QML	5%	0.20	0.03*	0*	0.01*	0.09	0.07	0.02*	0.42	0.30	0.75	0.05*	0.19	0.01*	6
	1%	0.51	0.18	0.34	0.14	0.31	0.19	0.93	0.09	0.82	0.19	0.25	−0.68	0.15	0
FHS-ST	5%	0.38	0.05*	0.01*	0.01*	0.11	0.01*	0.02*	0.40	0.39	0.64	0.02*	0.21	0.01*	7
	1%	0.43	0.15	0.09	0.06	0.02*	0.20	−0.95	0.07	0.88	0.15	0.25	−0.47	0.05	1
FHS-AST	5%	0.28	0.03*	0*	0.07	0.04*	0.01*	0.02*	0.39	0.40	0.80	0.08	0.25	0.01*	6
	1%	0.38	0.15	0.15	0.03*	0.11	0.10	0.80	0.05*	0.88	0.28	0.50	−0.52	0.03*	3
EVT-QML	5%	0.48	0.20	0*	0.03*	0.28	0.24	0.10	0.72	0.91	−0.69	0.05*	0.10	0*	4
	1%	0.24	0.07	0.03*	0.03*	0.29	0.41	0.11	0.05*	0.29	0.06	0.83	−0.45	0.23	3
EVT-ST	5%	0.56	0.23	0.02*	0.09	0.28	0.06	0.18	0.85	0.98	−0.58	0.05	0.08	0*	2
	1%	0.29	0.02*	0*	0.15	0.04*	0.28	0.30	0.13	0.34	0.07	0.58	−0.30	0.11	3
EVT-AST	5%	0.55	0.21	0.01*	0.15	0.12	0.07	0.18	0.91	0.98	−0.42	0.17	0.07	0*	2
	1%	0.40	0.03*	0.01*	0.03*	0.08	0.25	0.25	0.18	0.32	0.06	0.80	−0.61	0.09	3
REVT-QML	5%	−0.34	−0*	−0.71	0.06	−0.01*	0.04*	−0*	−0*	−0*	−0*	−0.48	0.01*	0*	9
	1%	0.01*	−0*	0*	0.08	0.56	0.11	−0.66	0.90	−0.13	0.20	−0.44	0*	0*	5
REVT-ST	5%	−0.23	−0*	−0.01*	0.04*	−0*	0.02*	−0*	−0*	−0*	−0*	−0.22	0*	0.11	10
	1%	0.10	−0*	0.11	0.02*	0.47	0*	−0.67	−0.98	−0.10	0.05	−0.65	0*	0*	5
REVT-AST	5%	−0.25	−0*	−0*	−0.66	−0*	0.38	−0*	−0*	−0*	−0*	−0.02*	0.03*	0.87	9
	1%	0.01*	−0*	0.11	0.03*	0.24	0.01*	−0.69	0.54	−0.38	0.07	−0.11	0*	0.04*	6
REVT-ROB	5%	0.54	0.68	−0*	0*	0.83	0.60	−0*	−0*	−0*	−0*	−0.68	0.05*	−0*	8
	1%	0.02*	0.01*	−0.09	0.01*	0.02*	0.04*	−0.74	0.33	−0.24	0.67	−0.80	0*	−0.77	6
NR		7	15	18	15	10	10	10	10	7	7	9	13	16	

Table 2.12: Expected Shortfall Mean Zero Test (5-Day Horizon)

This table lists the p-values for the expected shortfall mean zero test for the 5-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The considered threshold levels p, defining the level below which p% of the realized returns may fall, are 5% and 1%. In order to distinguish under- and over-estimations of risk, the p-values for which the forecasts are too conservative (over-estimation) carry a negative sign. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing model, showing the fewest rejections, is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	P-lev	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	5%	0.14	0.35	−0.34	0.02*	0.02*	0.01*	0.20	0.17	0.97	0.17	0*	0.94	−0.58	4
	1%	0.22	0.01*	0.20	0.13	0.72	0.18	0.13	−0.25	0.09	−0.31	0.06	0.45	0.67	1
FP-T	5%	0.98	1.00	−0.06	0.96	0.15	0.08	0.61	0.21	−0.63	0.64	0.04*	−0.02*	−0.01*	3
	1%	0.89	0.01*	0.28	0.63	0.41	0.03*	0.45	−0.13	0.15	−0.01*	0.19	−0.31	0.26	3
FP-ST	5%	−0.27	0.62	−0.33	−0.92	0.50	0.35	0.54	0.22	0.24	0.34	0.22	−0.03*	−0.04*	2
	1%	0.56	0.14	0*	0.47	−0.20	0.31	0.73	−0.10	0.24	−0.01*	0.23	−0*	0.52	3
FP-AST	5%	−0*	0.63	−0.41	−0.34	0.64	−0.96	0.44	0.19	−0.55	−0.30	0.22	−0*	−0.01*	3
	1%	0.59	0.01*	0.18	0.21	−0.12	0.61	0.63	−0.15	0.36	−0*	−0.69	−0*	0*	4
FHS-QML	5%	−0.40	0.45	−0.74	0.42	0.05	0.02*	0.43	0.27	−0.65	0.59	0.08	−0.19	−0.77	1
	1%	0.55	0.03*	0.62	0.19	−0.59	0.71	0.45	−0.05	0.24	−0*	0.23	−0.04*	0.48	3
FHS-ST	5%	−0.72	0.70	0.90	0.07	0.02*	0.01*	0.29	0.17	−0.48	0.43	0.03*	−0.29	−0.84	3
	1%	−0.62	0.09	0.18	0.02*	0.89	0.16	0.74	−0.08	0.35	−0.18	0.35	−0.63	0.57	1
FHS-AST	5%	0.92	0.52	−0.91	0.25	0.04*	0.02*	0.21	0.38	−0.50	0.73	0.02*	−0.31	−0.66	3
	1%	−0.69	0.01*	0.05	0.11	−0.76	0.55	0.35	−0.14	0.30	−0.74	0.32	−0.23	0.28	1
EVT-QML	5%	−0.74	0.48	0.83	0.32	0.07	0.04*	0.38	0.32	−0.72	0.58	0.05*	−0.31	−0.63	2
	1%	0.89	0.03*	0.48	0.23	−0.78	0.48	0.78	−0.25	0.14	−0.19	0.46	−0.64	0.39	1
EVT-ST	5%	−0.91	0.60	0.96	0.14	0.02*	0*	0.09	0.26	−0.95	0.20	0.03*	−0.46	−0.72	3
	1%	−0.86	0.03*	0.16	0.09	0.78	0.08	0.43	−0.21	0.17	−0.50	0.53	0.34	0.36	1
EVT-AST	5%	0.90	0.44	0.94	0.24	0.02*	0.01*	0.35	0.17	−0.87	0.92	0.02*	−0.36	−0.53	3
	1%	0.99	0.02*	0.23	0.71	−0.38	0.24	0.36	−0.20	0.16	−0.49	0.50	−0.92	0.28	1
REVT-QML	5%	−0.40	0.66	−0.57	0.29	0.19	0.04*	0.87	0.94	−0.34	0.79	0.09	−0.27	−0.59	1
	1%	−0.54	0.08	−0.59	0.22	−0.15	0.44	0.89	−0*	0.16	−0.11	−0.85	0.44	0.37	1
REVT-ST	5%	−0.60	0.90	−0.55	0.11	0.11	0*	0.42	0.89	−0.53	0.68	0.02*	−0.50	−0.48	2
	1%	−0.62	0.08	0.70	0.10	−0.12	0.06	0.67	−0*	0.26	−0.59	0.86	0.15	−0.51	1
REVT-AST	5%	−0.83	0.78	−0.62	0.23	0.11	0.02*	0.61	0.41	−0.53	−0.77	0.07	−0.39	−0.32	1
	1%	−0.81	0.06	0.67	0.77	−0*	0.23	0.99	−0*	0.28	−0.38	−0.73	−0.79	0.84	2
REVT-ROB	5%	0.88	0.72	0.82	0.15	0.32	0.41	0.50	0.64	0.96	−0.13	0.09	−0.11	−0.19	0
	1%	0.78	0.21	0.91	0.17	−0.24	0.71	−0.79	−0*	0.71	−0.55	0.92	−0.90	0.57	1
NR		1	8	1	2	6	11	0	4	0	4	8	6	4	

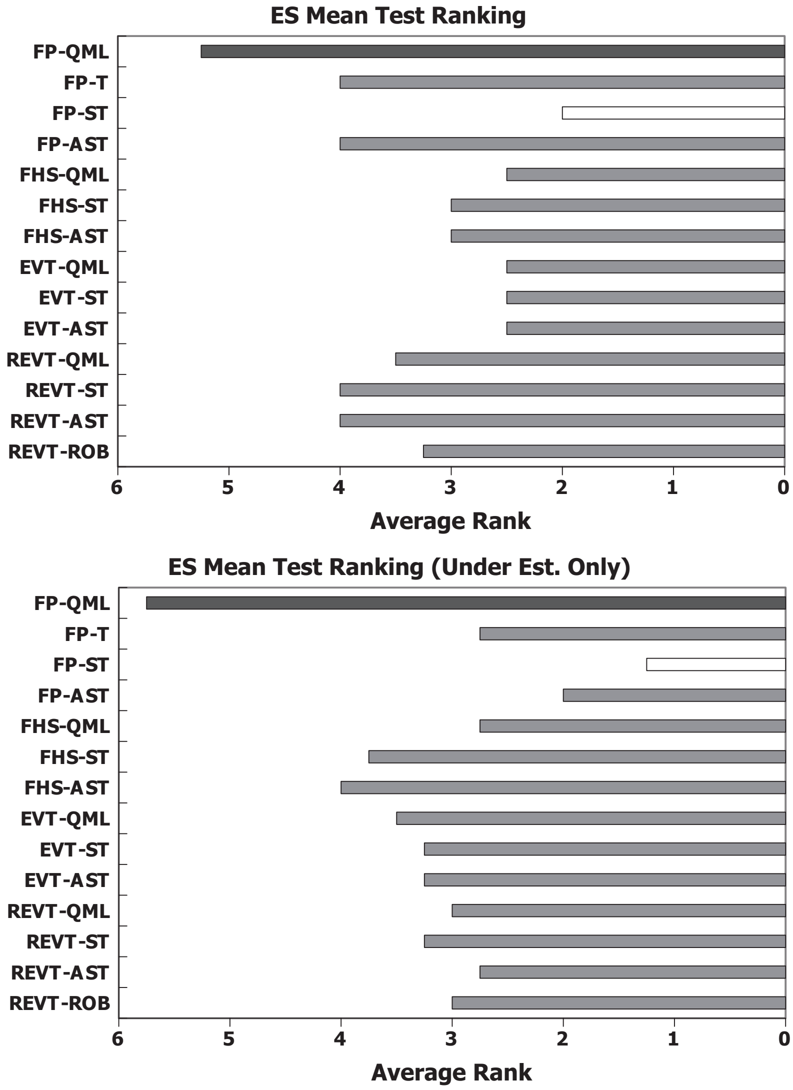
Table 2.13: Summary Expected Shortfall Test

This table summarizes the results for the expected shortfall mean zero test by listing and aggregating the number of rejections (NR) at the 5% level of significance for each threshold level (5% and 1%) as well as for both considered forecasting horizons (1d and 5d). Thereby, the number of rejections due to an under- respectively an over-estimation of risk are separately shown for each threshold level and forecasting horizon combination. The best performing model, showing the fewest rejections overall (column All Total), is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

	5% 1d Under	5% 1d Over	5% 1d Total	1% 1d Under	1% 1d Over	1% 1d Total	All 1d Total	5% 5d Under	5% 5d Over	5% 5d Total	1% 5d Under	1% 5d Over	1% 5d Total	All 5d Total	All Total	All Under	All Over
FP-QML	13	0	13	13	0	13	26	4	0	4	1	0	1	5	31	31	0
FP-T	4	2	6	0	6	6	12	1	2	3	2	1	3	6	18	7	11
FP-ST	0	0	0	0	0	0	0	0	2	2	1	2	3	5	5	1	4
FP-AST	0	6	6	2	3	5	11	0	3	3	2	2	4	7	18	4	14
FHS-QML	6	0	6	0	0	0	6	1	0	1	1	2	3	4	10	8	2
FHS-ST	7	0	7	1	0	1	8	3	0	3	1	0	1	4	12	12	0
FHS-AST	6	0	6	3	0	3	9	3	0	3	1	0	1	4	13	13	0
EVT-QML	4	0	4	3	0	3	7	2	0	2	1	0	1	3	10	10	0
EVT-ST	2	0	2	3	0	3	5	3	0	3	1	0	1	4	9	9	0
EVT-AST	2	0	2	3	0	3	5	3	0	3	1	0	1	4	9	9	0
REVT-QML	3	6	9	4	1	5	14	1	0	1	0	1	1	2	16	8	8
REVT-ST	3	7	10	4	1	5	15	2	0	2	0	1	1	3	18	9	9
REVT-AST	1	8	9	5	1	6	15	1	0	1	0	2	2	3	18	7	11
REVT-ROB	2	6	8	6	0	6	14	0	0	0	0	1	1	1	15	8	7

Figure 2.5: Ranking Expected Shortfall Mean Zero Test

The upper graph shows the results of applying a ranking procedure to the results summarized in table 2.13 in the columns labeled *Total*. The procedure works as follows: For every threshold level (5% or 1%) and every forecasting horizon (1 day or 5 days) the models are ranked according to their performance from the smallest number of rejections to the largest number of rejections. In a second step, the mean rank achieved by all models is calculated which is displayed in the upper graph. The best performing model, with the lowest rank, is indicated by a white bar, while the weakest performing model is indicated by a dark grey bar. The bottom graph follows the same procedure, but uses the under-estimation columns labeled *Under* in the summary table.



Similar to subsection 2.4.1, the performance of the different risk models with regards to the ES estimation is also being analyzed from an economic point of view. The setting is similar to the one discussed in subsection 2.4.1, however instead of considering the days between VaR violations, the Dollar amount losses which occur once a VaR violation occurs are of interest. In this setting the risk manager uses the risk model of choice in such a way, that the exposure in each index is adjusted on a daily basis, such that an ES, at the given threshold level, of 1 million (MM) USD is obtained. If the model correctly estimates the ES, the average loss when a VaR violation occurs is equal to 1 MM USD. When this procedure is applied to every considered index and model in this study, a box plot of the actual realized losses for every model can be plotted as done in figure 2.6[24]. Again, the focus will be on 1-day forecasts for the threshold levels 1% (upper graph) and 5% (bottom graph). As is seen for the 1% threshold level (upper graph), the median shortfall per VaR violation is under-estimated for all models except the FP-T and FP-AST models which over-estimate it. The weakest performing model is the FP-QML model which underestimates the shortfall for every index under consideration[25]. The median shortfall for the FP-QML model is 1.15 MM USD, and the mean shortfall is 1.145 MM USD. Thus, when using the QML model in the given setup, the average shortfall turns out to be almost 15% higher than predicted by the model. The best performing model FP-ST on the other side provided for a median shortfall of 1.01 MM USD and an average shortfall of 1.001 MM USD, thus complying very closely with the stated goal of the risk manager. Further, the deviation from the expected shortfall for the different indices on average is very low in the case of the FP-ST model. Moving to the 5% threshold level (bottom graph), the results are very similar, with the FP-QML model showing the weakest performance and the FP-ST model showing the strongest performance.

The analysis in figure 2.6 gives an idea of how the models perform once a shortfall occurs, however it does not take into account how often these

[24]A practical way to view the setup is a risk manager who oversees 13 portfolio managers each responsible of tracking one of the considered indices in this study. Per VaR violation every portfolio manager is allowed to only loose 1MM USD on average. In order to achieve this goal, the risk manager uses a risk model in order to determine how much risk needs to be hedged each day for every portfolio manager, such that the risk of the portfolio corresponds to the expected shortfall of 1MM USD.

[25]This is seen by the lower line extension of the box plot for the FP-QML model not reaching the expected shortfall line.

shortfalls occur. In order to also take into account the accuracy of the models at predicting the number of VaR violations, as investigated in figure 2.4, an analysis combining the results of figure 2.4 and 2.6 is being conducted. The idea is simply to calculate the average expected shortfall experienced over a 100 respectively 20-day period, corresponding to the 1% and 5% violation thresholds. If the model under consideration correctly estimates the VaR as well as the ES, the average shortfall over 100 respectively 20 days will equal 1 MM USD. The numbers for each model in combination with every index are retrieved by summing-up the shortfalls derived for the analysis in figure 2.6 and dividing them by the expected number of days between VaR violations. The results for this analysis are shown in figure 2.7. With regards to the 1% threshold (upper graph) it can be seen that the gap between the best performing and worst performing model has now increased considerably compared to the previous case study results. The median average shortfall over a 100 day period stands at 2.04 MM USD, more than double the expected 1 MM USD. The best performing models are the FP-ST and FP-AST with median average shortfalls of 1.1 MM USD respectively 0.95 MM USD. Also, the variation among the different indices are relatively narrow for the two best performing models. This comparison gives an idea of the economic costs of using an imprecise risk model for the given set of indices. The FHS and EVT based models all under-estimate the risk while the REVT models show a very wide dispersion among the results for the different indices. Continuing with the 5% threshold (bottom graph) it can be seen that all models generally perform better and that the differences between the models become smaller. The worst performing model is still the FP-QML model, while the REVT models now perform the best, closely followed by the FP-ST and FP-AST models. The FHS and EVT models are again inferior to the FP-ST and FP-AST respectively the REVT models. Given the result that the REVT models showed a wider dispersion among the estimates for the single indices at the 1% threshold level, from the risk manager's perspective the FP-ST and FP-AST models seem to be the preferred choice among all risk models from an economic point of view. The next subsection is focusing on the goodness-of-fit of the tail distribution prediction.

Figure 2.6: Case Study - Average Shortfall Per VaR Violation
The graphs in this figure depict box-plot diagrams of the average shortfall per VaR violation for each model when applied to the 13 indices considered in this study. Thereby, it is assumed that the exposure to every index is hedged on a daily basis in such a way that an expected shortfall of 1MM USD is expected for every index at the corresponding threshold level. The upper graph is related to the 1% threshold level, and the bottom graph is related to the 5% threshold level.

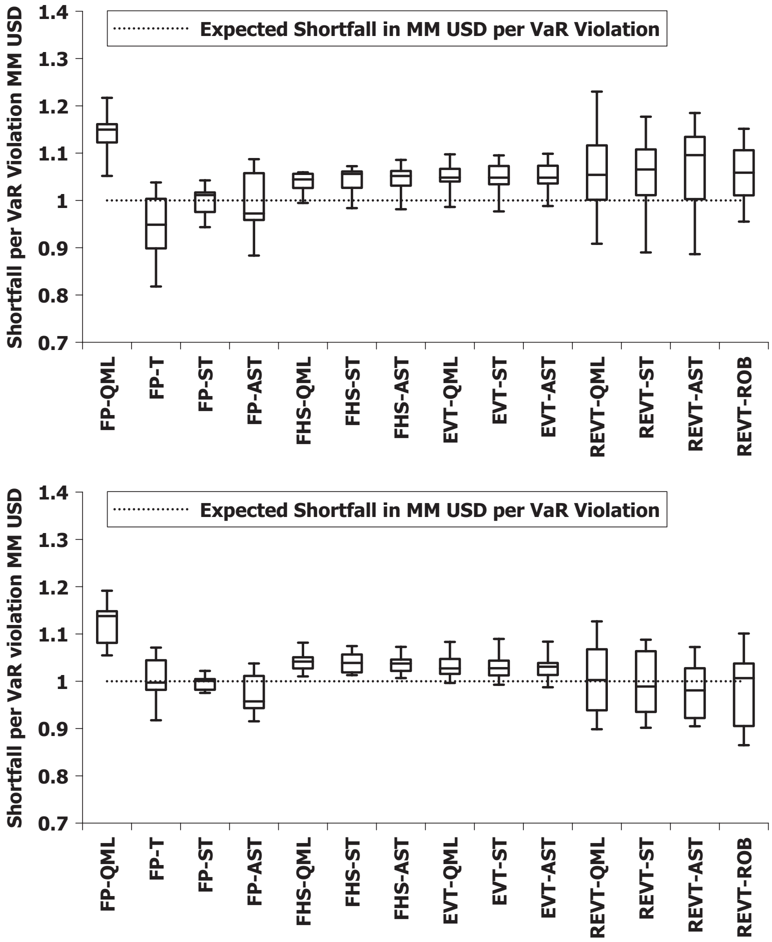
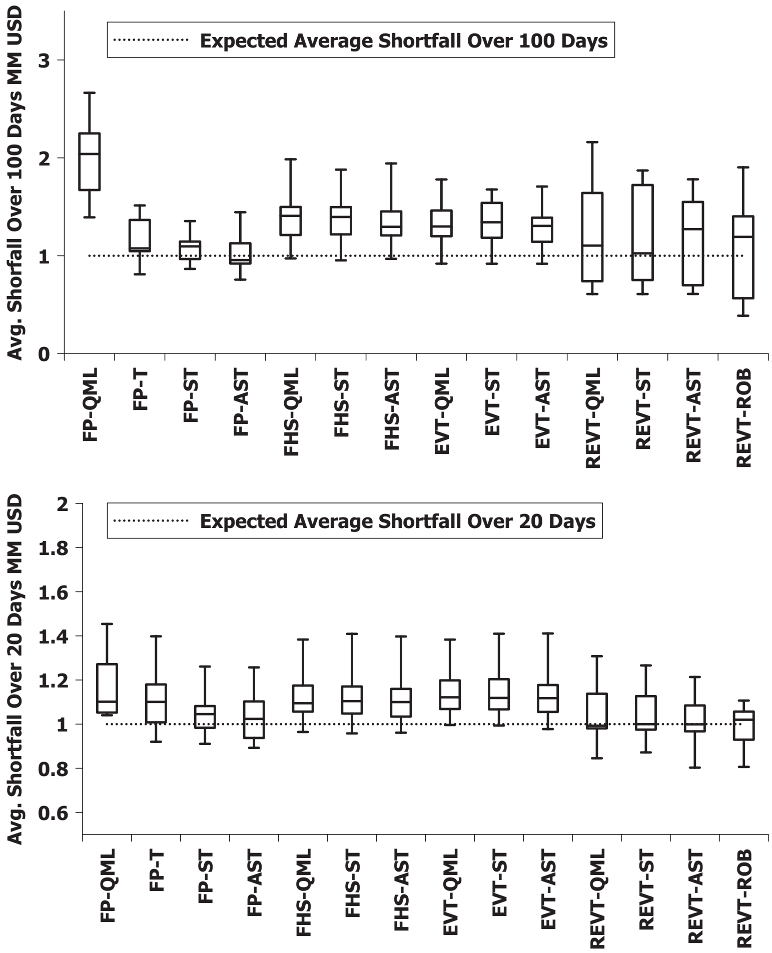


Figure 2.7: Case Study - Average Shortfall Over 100 and 20 Days

The graphs in this figure depict box-plot diagrams of the average shortfall over 100 respectively 20 days for each model when applied to the 13 indices considered in this study. Thereby, it is assumed that the exposure to every index is hedged on a daily basis in such a way that an expected shortfall of 1MM USD is expected for every index at the corresponding threshold level. The upper graph is related to the 1% threshold level, and the bottom graph is related to the 5% threshold level.



2.4.3 Testing the Goodness-of-Fit of Tail Distribution Predictions

For testing the goodness of fit of the tail distribution predictions, the goodness-of-fit test procedure described in subsection 2.3.2 is used. The following tables hold the p-values of the Pearson-Chi-Squared tests for the null hypothesis that the transformed returns in the tail correspond to a discrete uniform distribution. The suffix * highlights p-values for which the null hypothesis can be rejected at the 5% level of significance.

Analyzing the results for the 1-day horizon in table 2.14, with the summary found in table 2.16, the best performing models overall are FP-ST, FP-T and FP-AST with zero rejections (column: *1d*). The worst performing model is FP-QML with 11 rejections. With regards to the indices, EVAL, CRED, CUMO and CUVA are showing the highest number of rejections with 6 each as seen in the bottom row of table 2.14. Similar to the previous tests, the differences between the models become smaller when analyzing the 5-day horizon for which the individual results are shown in table 2.15. The best model is FP-ST with 0 rejections followed by FP-AST, FHS-AST, REVT-QML, REVT-AST and REVT-ROB with 1 rejection each (column: *5d*). The weakest model is FP-QML with 5 rejections. With regards to the indices, CUCA has the highest number of rejections counting 10 rejections of the null hypothesis at the 5% level of significance. Overall, as seen in the last column of table 2.16, the best performing model is FP-ST with 0 rejections followed by the FP-AST with 1 rejection. Similar to the tests for the VaR forecasting and the ES forecasting, the robust estimation procedures are unable to outperform the more sophisticated fully parametric estimation procedures (FP-ST and FP-AST).

Table 2.14: Goodness-of-Fit Tail Distribution Test (1-Day Horizon)

This table lists the p-values for the goodness-of-fit tail distribution test for the 1-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing models, showing the fewest rejections, are highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	0*	0*	0*	0*	0*	0*	0*	0.21	0.21	0*	0*	0*	0*	11
FP-T	0.80	0.59	0.23	0.86	0.74	0.72	0.28	0.13	0.74	0.21	0.08	0.13	0.82	0
FP-ST	0.95	0.96	0.90	0.77	0.77	0.68	0.33	0.66	0.93	0.08	0.35	0.75	0.25	0
FP-AST	0.43	0.36	0.78	0.52	0.46	0.84	0.28	0.93	0.90	0.11	0.53	0.65	0.35	0
FHS-QML	0.39	0.50	0*	0.05*	0.06	0.15	0.06	0.55	0.12	0.77	0.25	0.49	0.08	2
FHS-ST	0.45	0.62	0.02*	0.23	0.48	0.02*	0.06	0.65	0.48	0.69	0.30	0.11	0.14	2
FHS-AST	0.50	0.45	0.01*	0.29	0.31	0.05	0.12	0.65	0.56	0.95	0.33	0.16	0.28	1
EVT-QML	0.84	0.59	0.10	0.27	0.51	0.07	0.28	0.38	0.85	0.18	0.02*	0.18	0*	2
EVT-ST	0.73	0.83	0.15	0.20	0.87	0.01*	0.65	0.56	0.93	0.20	0.22	0.06	0.01*	2
EVT-AST	0.87	0.75	0.68	0.88	0.67	0.01*	0.64	0.47	0.94	0.13	0.50	0.02*	0.01*	3
REVT-QML	0.80	0.14	0.01*	0*	0.80	0.01*	0.01*	0.07	0.03*	0.01*	0.52	0*	0*	8
REVT-ST	0.43	0.13	0.44	0.01*	0.41	0.01*	0.01*	0.01*	0.03*	0.01*	0.76	0*	0*	8
REVT-AST	0.09	0.06	0.07	0.33	0.85	0.06	0.01*	0.03*	0.02*	0*	0.83	0*	0.07	5
REVT-ROB	0.31	0.76	0*	0*	0.14	0.44	0*	0.11	0*	0*	0.12	0*	0.59	6
NR	1	1	6	5	1	6	5	2	4	5	2	6	6	

Table 2.15: Goodness-of-Fit Tail Distribution Test (5-Day Horizon)

This table lists the p-values for the goodness-of-fit tail distribution test for the 5-day horizon. The suffix * highlights p-values for which the null hypothesis is rejected at the 5% level of significance. The last column and the bottom row denoted by NR list the number of rejections at the 5% level of significance for the models respectively the indices under consideration. The best performing model, showing the fewest rejections, is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

Model	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA	NR
FP-QML	0.12	0.96	0.65	0.04*	0*	0.09	0.28	0.03*	0.11	0.02*	0*	0.16	0.78	5
FP-T	0.39	0.19	0.11	0.32	0.29	0.19	0.23	0.24	0.03*	0.09	0.05*	0.23	0.04*	3
FP-ST	0.65	0.08	0.09	0.11	0.44	0.74	0.68	0.14	0.55	0.13	0.34	0.14	0.22	0
FP-AST	0.42	0.08	0.23	0.81	0.43	0.60	0.28	0.08	0.52	0.36	0.01*	0.32	0.09	1
FHS-QML	0.72	0.60	0.31	0.14	0.02*	0.05*	0.17	0.09	0.62	0.32	0.04*	0.62	0.74	3
FHS-ST	0.48	0.05*	0.39	0.05	0.10	0.05*	0.29	0.08	0.28	0.35	0.02*	0.47	0.27	3
FHS-AST	0.10	0.44	0.35	0.19	0.07	0.02*	0.16	0.20	0.63	0.47	0.08	0.73	0.59	1
EVT-QML	0.19	0.90	0.33	0.22	0.14	0.09	0.41	0.08	0.24	0.55	0.02*	0.04*	0.97	2
EVT-ST	0.66	0.14	0.51	0.17	0.19	0.03*	0.06	0.25	0.22	0.35	0.03*	0.91	0.17	2
EVT-AST	0.51	0.15	0.83	0.06	0.04*	0.03*	0.36	0.16	0.53	0.82	0.06	0.93	0.41	2
REVT-QML	0.74	0.63	0.76	0.22	0.32	0.19	0.55	0.33	0.18	0.78	0.04*	0.41	0.99	1
REVT-ST	0.60	0.23	0.72	0.14	0.05*	0.02*	0.05	0.19	0.24	0.73	0.01*	0.81	0.18	3
REVT-AST	0.34	0.08	0.56	0.12	0.01*	0.12	0.14	0.06	0.28	0.91	0.09	0.96	0.38	1
REVT-ROB	0.27	0.19	0.64	0.26	0.16	0.54	0.55	0.22	0.33	0.44	0.05*	0.80	0.36	1
NR	0	1	0	1	5	6	0	1	1	1	10	1	1	

Table 2.16: Summary Goodness-of-Fit Tail Distribution Test

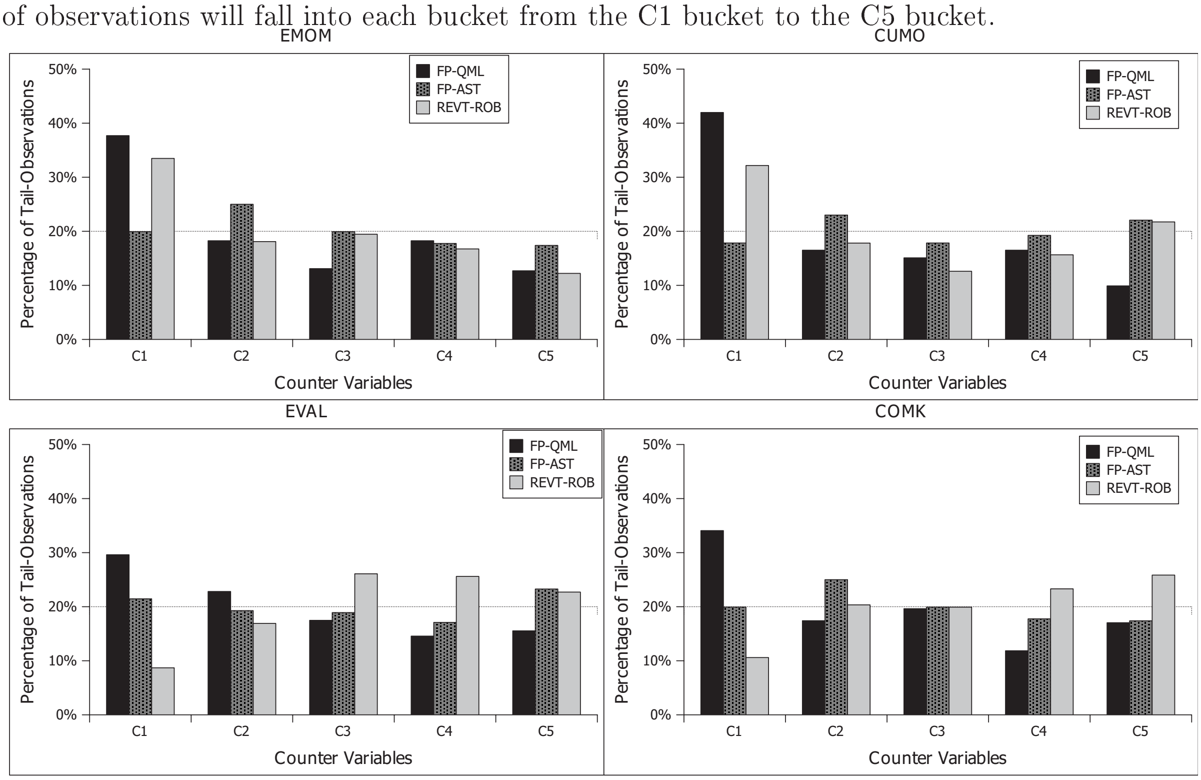
This table summarizes the results for the goodness-of-fit tail distribution test by listing and aggregating the number of rejections (NR) at the 5% level of significance for each threshold level (5% and 1%) as well as for both considered forecasting horizons (1d and 5d). The best performing model, showing the fewest rejections overall (column *All*), is highlighted in bold and the weakest performing model, showing the highest number of rejections, is highlighted in italic.

	1d	5d	All
FP-QML	11	5	16
FP-T	0	3	3
FP-ST	0	0	0
FP-AST	0	1	1
FHS-QML	2	3	5
FHS-ST	2	3	5
FHS-AST	1	1	2
EVT-QML	2	2	4
EVT-ST	2	2	4
EVT-AST	3	2	5
REVT-QML	8	1	9
REVT-ST	8	3	11
REVT-AST	5	1	6
REVT-ROB	6	1	7

As was already mentioned in the section on the VaR forecasting, it seems that the under-performance of the robust models can not be attributed solely to either over- or under-estimating the risk in the tail of the distribution. Rather, the robust methods tend to over- and under-estimate the distribution in the extreme tail, depending on the index under consideration. Further evidence for this phenomenon can be found by the visual inspection of the graphs in figure 2.8. The graphs show the transformed tail distribution of the realized returns as is is used to conduct the goodness of fit test as discussed in section 2.3.2 for the 1-day horizon for several indices. If a model correctly predicts the distribution of the tail on average, the same number of observations will fall into each bucket from C1, corresponding to the]-Inf 1%] range quantile, to the C5 bucket, corresponding to the]4%, 5%] range quantile. Thus, every bucket should hold approximately 20% of the observations. As is seen in the upper two graphs, the FP-QML and REVT-

ROB models tend to under-estimate the risk in the far tail for the EMOM and CUMO indices, since significantly more returns are observed in bucket C1 than the 20% which are expected when the model is in fact correctly predicting the tail. However, as seen in the lower two graphs in figure 2.8, for the EVAL and COMK indices, the REVT-ROB model over-estimates the risk since too few observations fall into the C1 bucket and too many observations are found in the other buckets. For the EVAL and COMK indices, the FP-QML model still under-estimates the risk while the FP-AST model performs relatively well at describing the tail of the distribution in all 4 cases.

Figure 2.8: Transformed Tail Distribution Comparison
The graphs show the transformed tail distributions of the realized returns as they are used to conduct the goodness of fit test as discussed in section 2.3.2 for the models FP-QML, FP-AST and REVT-ROB applied to the indices EMOM, CUMO, EVAL and COMK for the 1-day horizon. If a model correctly predicts the distribution of the tail on average, the same percentage of observations will fall into each bucket from the C1 bucket to the C5 bucket.



2.5 Conclusion

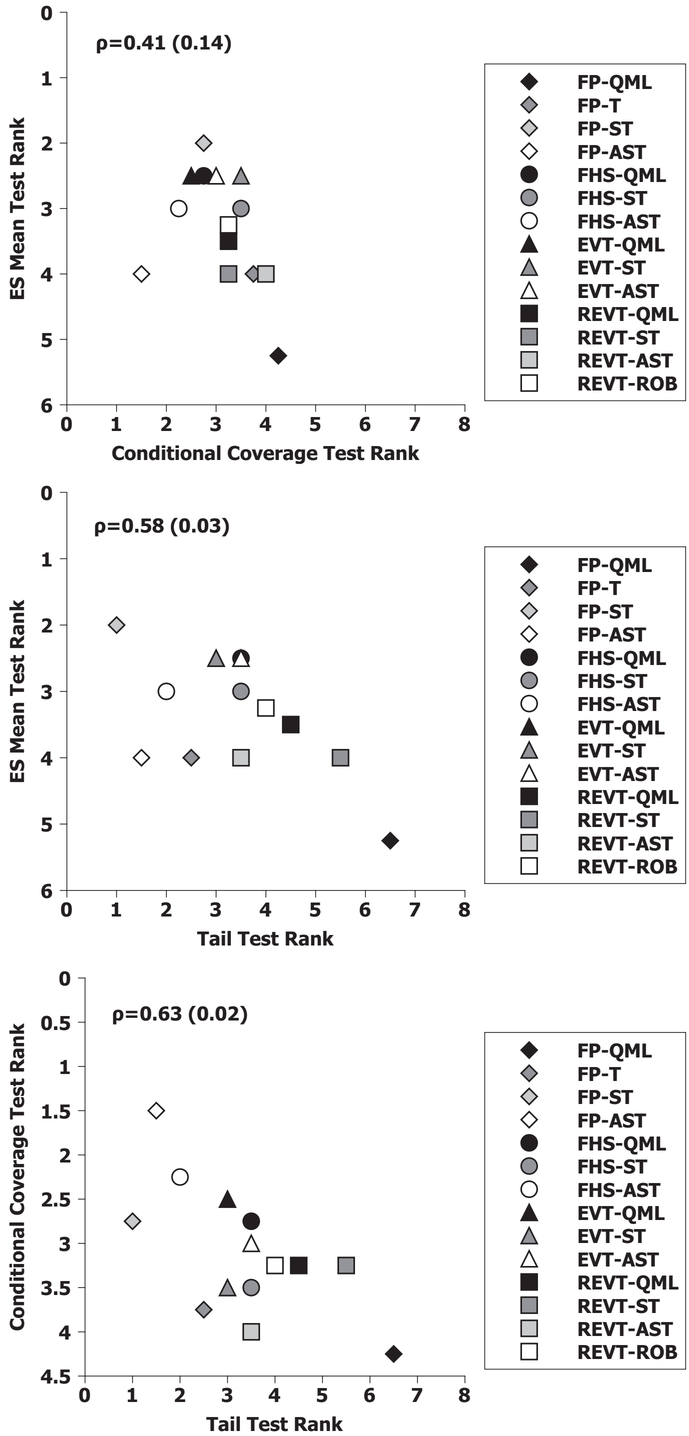
This study focuses on testing the performance of a broad set of risk models on various indices and a wider set of risk metrics in comparison to most previous studies. With regards to the indices, this study tests the risk models not only with respect to different asset classes, but also with respect to different trading strategies. The results of the study indicate, that the considered models do not work equally well across all markets and trading strategies. However, applying the different models to the dynamic trading strategies did not give any indication that the considered models are generally less suited to describe the risks of these strategies compared to static indices. With regards to the tested risk metrics, this study put more emphasis on ES as well as the goodness-of-fit for the prediction of the tail distribution instead of solely focusing on VaR. Further, this study has included some newer risk models which have recently been proposed in the literature and compared the performance to well established approaches.

The results of the study provide evidence that in the context of the proposed indices, some models do indeed perform better on average than others. With regards to the VaR related tests considered in subsection 2.4.1, a variant of the fully parametric model using a generalized asymmetric student-t distribution as proposed in Zhu and Galbraith (2010) and first applied in Zhu and Galbraith (2011) (FP-AST) showed the best performance across the tested indices, threshold levels and time horizons. Also, when only considering the under-estimation of VaR, as done for the unconditional coverage test, the FP-AST model is among the best performing models. An interesting result with respect to the VaR related tests is the performance of the models relying on robust estimation procedures as proposed in Mancini and Trojani (2011). When analyzing the 1-day forecasting horizon, these models have performed very well with respect to forecasting VaR at the 5% threshold when considering the unconditional coverage test. However, at the 1% threshold the performance deteriorated and could not compete with the other advanced, non-robust approaches. Nevertheless, when only considering the test for unconditional coverage, the robust REVT-ROB model is the second best performing model overall. Further, the robust models were not able to compete with the best performing fully parametric models when applying the independence test at the 1-day horizon. The results for the VaR related test were corroborated by the case study analysis which compared the different models from an economic point of view.

The results for the ES mean zero test indicated that it is crucial whether a one-sided test or a two-sided test is applied. When using a one-sided test, as done for example in McNeil and Frey (2000), the parametric models FP-ST and FP-AST are showing the strongest performance. When a two-sided test is applied, taking into account a possible over-estimation of risk, the FP-ST model followed by the EVT models and the FHS-QML model performed best. Again, the models based on robust estimation procedures were not able to compete with the best performing non-robust models. The case study analysis conducted with respect to the average shortfall per VaR violation provided insights into the economic costs of using an inaccurate risk model versus an accurate risk model. Further, the case study analyzing the average shortfall over defined time periods, thereby combining the results from the VaR and ES analysis, revealed that the choice of the risk model may have a vast impact from a financial and economic point of view. With regards to the case study, the FP-AST as well as the FP-ST models looked the most promising from a risk manager's point of view. For the goodness-of-fit test conducted in subsection 2.4.3, it was again the FP-ST model followed by the FP-AST model which showed the strongest performance.

Given these results, the question rises whether a best overall model can be identified. As can be seen in figure 2.9, the rankings of the models obtained for the different tests in this study are correlated to some extent. The correlation coefficients and the corresponding p-values in parentheses are shown in the graphs. This means that models which perform well with respect to one test tend to perform well at the other tests as well.

Figure 2.9: Correlation Across Rankings
The graphs plot pairwise the ranking results for tested models of the conditional coverage test (upper graph in figure 2.3), the ES mean test (upper graph in figure 2.5) and the ranks obtained by ranking the models based on the results found in the last column of table 2.16 (referring to the goodness-of-fit tail distribution test). The correlation between the ranks is denoted by ρ in each graph and the corresponding p-value is given in parentheses.



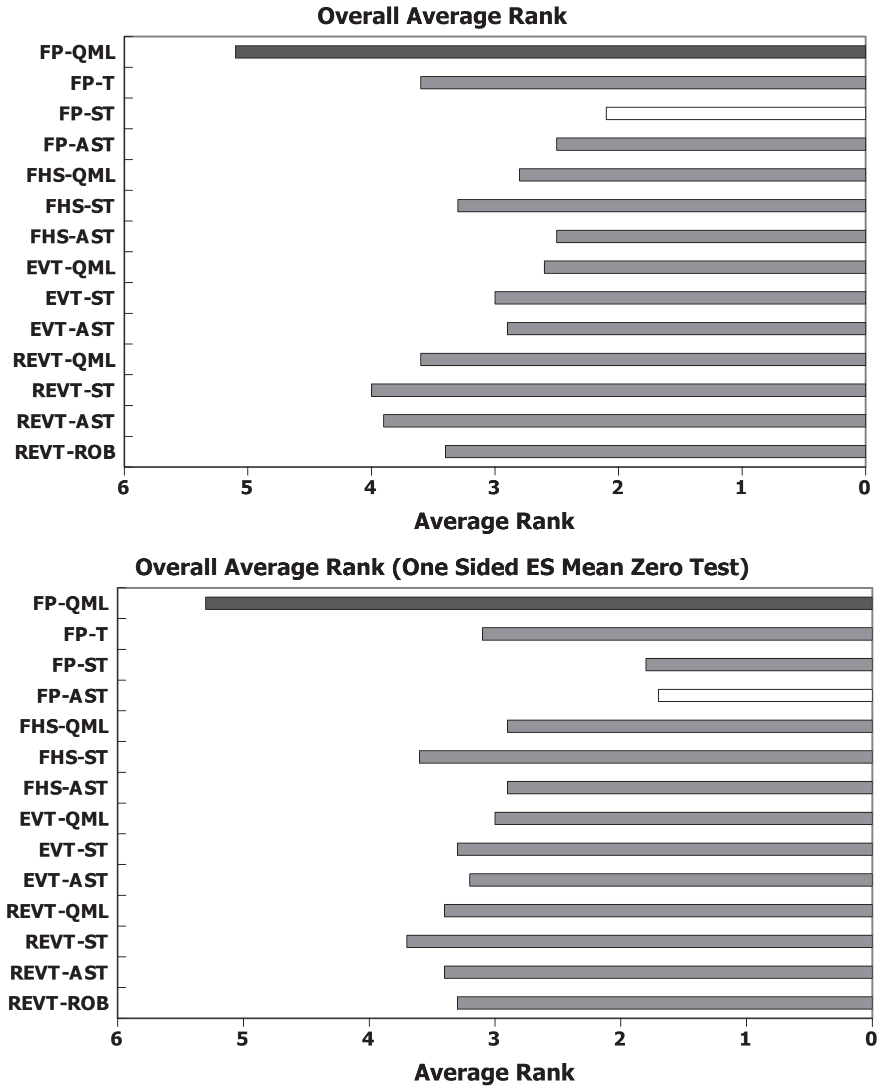
In order to give guidance in determining the best performing model, a ranking procedure is applied as done for various tests in section 2.4. The procedure ranks the performance of the models in the conditional coverage test as given in table 2.10, the ES mean zero test as given in table 2.13 and the goodness of fit test as given in table 2.16 for the different horizons and thresholds where appropriate. Then, the average rank is calculated as displayed in the upper graph in figure 2.10[26]. Also, a second version of the overall ranking is produced, by using the one-sided test results in the ES mean zero test, thus only accounting for an under-estimation of the ES[27]. This result is shown in the bottom graph of figure 2.10.

[26]In detail the procedure works as follows: For the summary of the conditional coverage test as given in table 2.10, the rank of each model for each of the following columns is determined: *5% 1-d*, *1% 1-d*, *5% 5-d*, and *1% 5-d*. Then, for the summary of the ES mean zero test, as given in table 2.13, the rank of each model for each of the following columns is determined: *5% 1-d Total*, *1% 1-d Total*, *5% 5-d Total*, and *1% 5-d Total*. Finally, for the summary of the goodness-of-fit tail distribution test given in table 2.16, the ranks of each model for the two columns *1d* and *5d* are determined. This procedure gives a total of 10 ranks for every model for which the average is calculated and displayed in figure 2.10.

[27]The ranks of the columns *5% 1-d Under*, *1% 1-d Under*, *5% 5-d Under*, and *1% 5-d Under* in table 2.13 are used instead of the columns *5% 1-d Total*, *1% 1-d Total*, *5% 5-d Total*, and *1% 5-d Total*.

Figure 2.10: Ranking Overall

The upper graph shows the results of applying a ranking procedure to the results in the conditional coverage test as given in table 2.10, the ES mean zero test as given in table 2.13 and the goodness of fit test as given in table 2.16 for the different horizons and thresholds where appropriate. The procedure works as follows: For the summary of the conditional coverage test as given in table 2.10, the rank of each model for each of the following columns is determined: *5% 1-d*, *1% 1-d*, *5% 5-d*, and *1% 5-d*. Then, for the summary of the ES mean zero test, as given in table 2.13, the rank of each model for each of the following columns is determined: *5% 1-d Total*, *1% 1-d Total*, *5% 5-d Total*, and *1% 5-d Total*. Finally, for the summary of the goodness-of-fit tail distribution test given in table 2.16, the ranks of each model for the two columns *1d* and *5d* are determined. This procedure gives a total of 10 ranks for every model for which the average is calculated The best performing model, with the lowest rank, is indicated by a white bar, while the weakest performing model is indicated by a dark grey bar. The bottom graph follows the same procedure, but uses the under-estimation columns labeled *Under* in the ES mean zero test table 2.13.



The results in figure 2.10 indicate that the FP-ST, followed by the FP-AST and FHS-AST models have shown the best performance in the given setting when a two-sided ES mean zero test is applied. When a one-sided test is used, the FP-AST model performs best, followed by the FP-ST, the FHS-AST and the FHS-QML model. Thus, the overall result reveal that from an overall performance perspective the more sophisticated fully parametric models are able to clearly outperform the other approaches in the given setting, including the robust approaches. This confirms the intuitive results from the economic case study. From the recently proposed models, the FP-AST approach introduced by Zhu and Galbraith (2011) seems to be worthwhile to be considered for risk modeling purposes. Notably, the results indicate that when a too conservative ES estimate is not a concern, the FP-AST model is the best performing model overall. The results of this study confirm the findings of Kuester, Mittnik and Paolella (2006), that the distributional assumptions in the filtering process are crucial for the outcome of the risk model performance. Further, the results of this study showed clear evidence that the choice of the risk model becomes less important at the 5-day forecasting horizon as all models tend to perform better at the longer horizon and the performance differences between the models become smaller. An explanation for this phenomenon is likely related to the distributional properties of the 5-day returns compared to the 1-day returns. As it is seen in table 2.19 in the appendix, the kurtosis of the index returns are generally lower compared to the corresponding 1-day returns for which the metrics are found in table 2.1. The heavier tails of the 1-day returns compared to the 5-day returns are also clearly observable for most indices in the QQ-plots in figure 2.11 in the appendix.

Further, the study found evidence, that the models based on robust estimation procedures were not able to compete with the non-robust models in the given setting. However, it needs to be acknowledged that there is some leeway with respect to the choice of parameter values in the estimation procedure which, for this study, are based on values used in previous studies. The sensitivity of the performance of the different risk models applied in this study with respect to the parameter specifications could be an interesting topic for future research. Also, additional research in the field of robust risk model estimation may be worthwhile. One interesting extension of the robust tail estimation procedure could for example exist in analyzing whether the introduction of an automated threshold selection procedure as proposed in Dupuis (1999) for determining the optimal value of u is worthwhile in the context of financial risk modeling.

2.6 Appendix

Commodity Momentum Strategy (COMO)

In order to create the commodity momentum factor (COMO), in a first step, investable indices for each commodity are downloaded from Bloomberg for the commodities listed in table 2.17. The single commodity indices are then used to construct a long-short momentum portfolio. I follow the approach of Miffre and Rallis (2007) by ranking the single commodity indices according to the performance in the ranking period which I choose to be 260 trading days. Thus, at the end of every month, the single commodity indices are ranked according to their 260 day performance from the best performing commodity to the worst performing commodity. Then, the long leg of the momentum portfolio is determined by giving equal (positive) weight to the top 7 single commodity indices in the ranking list and the short leg is determined by giving equal (negative) weight to the bottom 7 commodity indices in the ranking list[28]. The chosen holding period of the long-short portfolio is chosen to be 21 trading days after which the portfolio is being rebalanced by again ranking the commodity indices according to their 260 day performance.

[28] The weights assigned to the long leg sum to +1, and the weights assigned to the short leg sum up to -1. Thus the portfolio has a net exposure of zero.

Table 2.17: Commodity Indices

List of the commodity futures and the corresponding Bloomberg tickers used for the calculation of the COMO and COBA indices.

Commodity	Bloomber Ticker
WTI Crude	CL1
Brent Crude	CO1
Heating Oil	HO1
Gasoil	QS1
Natural Gas	NG1
Copper	HG1
Gold	GC1
Silver	SI1
Wheat	W 1
Corn	C 1
Soybean	S 1
Cotton	CT1
Sugar 11	SB1
Coffee	KC1
Cocoa	CC1
Live Cattle	LC1
Feeder Cattle	FC1
Lean Hogs	LH1
Orange Juice	JO1
Soybean Meal	SM1
Soybean Oil	BO1
Lumber	LB1

Commodity Backwardation Strategy (COBA)

Similar to the study of Fuertes, Miffre and Rallis (2010) when constructing the commodity backwardation strategy (COBA), I also determine first how much the available commodities are in backwardation respectively contango. Thereby, I examine the prices of the futures contract which settle in two months time which were also used for constructing the commodity momentum factor. I refer to this contract as the first contract and denote its price as P_1. In addition, I also use the price of the next contract which expires directly after the first contract and denote its price as P_2. The ratio $\frac{P1}{P2}$ determines the implied roll yield of each commodity. Similar to the commodity momentum approach, the commodities are ranked from the commodity with the highest ratio (strongest backwardation) to the one with the lowest ratio (strongest contango). The long leg of the portfolio is built by taking an equally weighted long position in the 7 most backwardated commodities and taking a short position in the 7 most contangoed commodities. The rebalancing frequency is chosen to be 21 days. The same set of commodities is being used as described in table 2.17.

Currency Carry Strategy (CUCA)

The carry trade strategy is implemented by first determining the implied 1-month interest rate differential for each currency i at time t by the following formula, where $F_{t,i}^{mid}$ and $S_{t,i}^{mid}$ refer to the mid 1-month forward rate and the mid spot rate at time t for currency i.[29]:

$$sigCarry_{t,i} = \frac{F_{t,i}^{mid}}{S_{t,i}^{mid}} - 1 \tag{2.18}$$

The used forward and spot foreign exchange rates are quoted in units of USD per one unit of foreign currency[30]. Therefore, if the interest rate differential

[29] The mid rates are calculated by taking the arithmetic average of the bid and ask prices.

[30] Thus, they represent the price a USD investor has to pay in order to buy one unit of foreign currency.

as determined by the above equation is <0, this implies that the foreign currency carries a higher interest rate than the USD and should therefore depreciate. In order to profit from the higher interest rate in the foreign currency and take a bet that it will depreciate by less than is implied by the interest rate differential at time t, an investor can enter into a long 1-month forward contract at the rate $F_{t,i}^{mid}$. After 1-month time, at $t+1$, the investor has to deliver dollars in the amount of $F_{t,i}^{mid}$ and receives one unit of foreign currency. The investor makes a profit on this trade whenever the spot rate $S_{t+1,i}^{mid}$, at which he can exchange the unit of foreign currency back into USD, is higher than the forward rate $F_{t,i}^{mid}$. The return $r_{t+1,i,long}$ in terms of the USD notional value invested in the forward contract at time t is therefore determined by:

$$r_{t+1,i,long} = \frac{S_{t+1,i}^{mid}}{F_{t,i}^{mid}} - 1 \tag{2.19}$$

Similarly, if the interest rate differential is >0 the investor would enter into a short 1-month forward contract and the return in terms of the USD notional value invested in the forward contract would equal to:

$$r_{t+1,i,short} = -\frac{S_{t+1,i}^{mid}}{F_{t,i}^{mid}} - 1 \tag{2.20}$$

The currency carry risk strategy CUCA is constructed by first determining the interest rate differential for the 12 currencies listed in table 2.18 versus the USD at the end of each month. Then, the long leg of the portfolio is defined by an equally weighted portfolio of long forward contracts of the 4 currencies for which the interest rate differential is the most negative. The short leg is created by an equally weighted portfolio of short forward contracts for the 4 currencies for which the interest rate differential is the most positive. The strategy is rebalanced monthly.

Table 2.18: Currencies Used for Strategy Calculations

List of the currency pairs and the corresponding Datastream tickers for the spot price and 1-month forward rates used for the calculation of the CUCA, CUMO and CUVA indices.

Currency Pair	Datastream Ticker Spot Price	Datastream Ticker 1-M Forward Price
USDAUD	TDAUDSP	TDAUD1F
CADUSD	TDCADSP	TDCAD1F
CHFUSD	TDCHFSP	TDCHF1F
DKKUSD	TDDKKSP	TDDKK1F
USDEUR	TDEURSP	TDEUR1F
USDGBP	TDGBPSP	TDGBP1F
HKDUSD	TDHKDSP	TDHKD1F
JPYUSD	TDJPYSP	TDJPY1F
NOKUSD	TDNOKSP	TDNOK1F
USDNZD	TDNZDSP	TDNZD1F
SEKUSD	TDSEKSP	TDSEK1F
SGDUSD	TDSGDSP	TDSGD1F

Currency Momentum Strategy (CUMO)

The currency momentum strategy CUMO and the currency value strategy CUVA are constructed in the same way as the currency carry strategy CUCA. The only difference is the definition of the signal which determines at the end of each month whether a long or short position is taken in the forward contracts. For the momentum strategy, the signal generation is simple. At the end of each month a momentum signal is generated which is equal to the spot rate return over the past three months:

$$sigMom_{t,i} = \frac{S_{t,i}^{mid}}{S_{t-3,i}^{mid}} - 1 \qquad (2.21)$$

The strategy assumption is, that currencies which have appreciated during the past 3-month, will continue to do so next month, while those which have depreciated will also continue on this path for the next month. Thus, as for the currency carry strategy, the long leg of the portfolio is constructed by an equally weighted portfolio of long forward contracts for the 4 currencies for which the positive momentum signal is the most positive. Similarly, the short leg of the portfolio is created by taking short positions in the 4 currencies for which the momentum signal is the most negative.

Currency Value Strategy (CUVA)

The last currency strategy is the value strategy. The idea behind this strategy, as put forward for example in Kroencke, Schindler and Schrimpf (2011), is that currencies trade above or below the exchange rate which is justified by the fundamental value of the currency as for example measured by the purchasing power parity (PPP). An investor exploiting a value strategy in currencies would therefore go long currencies which are undervalued and go short currencies which are overvalued from a fundamental point of view. In this study, I use the effective exchange rate (EER) indices published by the bank for international settlements (BIS) to determine the relative value of the currencies in my data set[31]. Similar to the previous two currency strategies, a signal is generated which indicates whether a currency is over- or undervalued against the USD and therefore determines whether a long or short position should be entered in the respective forward contracts. The signal is generated by taking the difference between the two month lagged effective exchange rate $EER_{t-2,i}$ of the foreign currency i and the lagged effective exchange rate of the USD $EER_{t-2,USD}$[32].

$$sigVal_{t,i} = EER_{t-2,i} - EER_{t-2,USD} \tag{2.22}$$

Again, as for the previous currency strategies, at the end of each month the long leg of the portfolio is being constructed by an equally weighted portfolio of long forward contracts for the 4 currencies which are the most undervalued. Similarly, an equally weighted portfolio of short forward contracts is defined by the 4 currencies which are most undervalued according to the

[31] The data can be downloaded under the following link:
http://www.bis.org/statistics/eer/index.htm

[32] The lag is introduced to ensure that only data is being used which could have been available at the time the portfolio is being rebalanced.

signal defined in equation 2.22.

The Idea of Robust Estimation Methods

As it is put forward in Mancini and Trojani (2011), the motivation for using robust estimators lies in the peril of a few data points having an overproportional effect on the estimation results making the estimates unreliable. Robust estimation procedures aim at limiting the potential damaging effects from outlying data points. As described by Wilcox (2012), there are different kinds of robustness by which estimators can be judged. For the sake of brevity I will only consider the so called infinitesimal robustness and follow closely the description and examples in Wilcox (2012) to demonstrate the basic idea as well as the linkage to the robust estimators used in this study.

To give an intuitive idea of the robustness concept, assume that an arbitrary function $f(x)$ is given and it needs to be determined which properties $f(x)$ should have in order for small changes in x not resulting in large changes of $f(x)$. As explained in Wilcox (2012), one such condition is that the function $f(x)$ is differentiable and that the derivative $f^{'}(x)$ is strictly smaller than a constant B. The condition $f(x) < B$ is referred to as boundedness of the derivative. As an example the function $f(x) = x^2$ does not have a bounded derivative as $2x$ increases without bound as x becomes large (Wilcox (2012)).

Instead of focusing on an arbitrary function $f(x)$, we can introduce a distribution F, and a functional $T(F)$, which represents for example a measure of location of the distribution F (for example the mean). In order for the chosen measure of location to be infinitesimal robust, the derivative of $T(F)$ needs to be bounded, whereby the derivative of the functional $T(F)$ is referred to as the influence function. As described by Wilcox (2012), the influence function can be interpreted as follows:

> Roughly , the influence function measures the relative extent a small perturbation in F has on $T(F)$. Put another way, it reflects

the (normed) limiting influence of adding one more observation, x, to a very large sample.

Thus, a way of achieving robustness is to choose a functional respectively a measure which has a bounded influence function, such that the influence on the result of any single observation in the data set is limited[33]. In order to illustrate more explicitly the mechanics of defining and estimating a robust statistic, the following example from Wilcox (2012) is utilized whereby an intuitive approach to so called M-estimators is given which are the kind of estimators used for the robust estimations in this study.

Let us assume that we try to find a measure of location, denoted by μ_m, which is close on average to all the possible values of the random variable X. A common way of quantifying how close a value μ_m is apart from all possible values of the random variable X, is to use the expected squared distance denoted by $E\left(X-\mu_m\right)^2$. In order to find the value μ_m which minimizes the expected squared distance, an optimization can be conducted by differentiating $E\left(X-\mu_m\right)^2$ with respect to μ_m and setting the expression equal to 0. This yields $-2E\left(X-\mu_m\right)=0$ which can be simplified into the following equation:

$$E(X-\mu_m)=0 \tag{2.23}$$

Solving equation 2.23 yields that μ_m is equal to the mean $\mu_m=\mu$. The problem of using the mean as a measure of location is that it is not robust since it exhibits an unbounded influence function as will be seen below. The way of obtaining a robust estimator of the location measure is to find an alternative function which measures the distance from a point and has desirable properties such as a bounded influence function. Translating the expression above from the special case of measuring the squared distance to a general case, the function $\xi(X-\mu_m)$ which measures the distance from

[33] The approach based on influence functions is also called infinitesimal approach and is described in detail in Hampel, Ronchetti and Rousseeuw (1986).

μ_m and its derivative $\psi(X - \mu_m)$ are introduced. By the same minimization procedure as explained for the special case of squared distances, the location measure μ_m is now determined by solving the following equation which is akin to equation 2.10:

$$E\left[\psi(X - \mu_m)\right] = 0 \tag{2.24}$$

The challenge consists of choosing ξ respectively ψ with desirable properties such as having a bounded influence function. Thereby, the influence function when μ_m is determined by equation 2.24 has the following form (Wilcox (2012)):

$$IF_m(x) = \frac{\psi(x - \mu_m)}{E\left[\psi'(X - \mu_m)\right]} \tag{2.25}$$

As can be seen from equation 2.25, when replacing $\psi(x - \mu_m)$ with the solution for the location measure based on the squared distance, which is $\psi(x - \mu_m) = x - \mu$, the influence function is not bounded as it increases indefinitely as x becomes larger. Various choices for ξ and ψ have been proposed in the literature whereby an overview can be found in Wilcox (2012).

The M-estimators relevant for this study are the robust M-estimator for GARCH type models as proposed in Mancini, Ronchetti and Trojani (2005) and the standardized optimal bias-robust estimator as used in Dupuis (1999) and defined in Hampel, Ronchetti and Rousseeuw (1986) and Victoria-Feser and Ronchetti (1994). Estimating the parameters of the M-estimators used in this study is conducted by solving the system of equations:

$$\sum_{j=1}^{G} \left[\psi_c\left(t_j, s\left(Y, \theta, \right)\right)\right] = 0$$

Thereby ψ_c is defined as in equation 2.11. The solution to the set of equations needs to be found iteratively and is shown for example in the appendix of Dupuis (1999). The estimation is computationally demanding as numerical integrations need to be deployed.

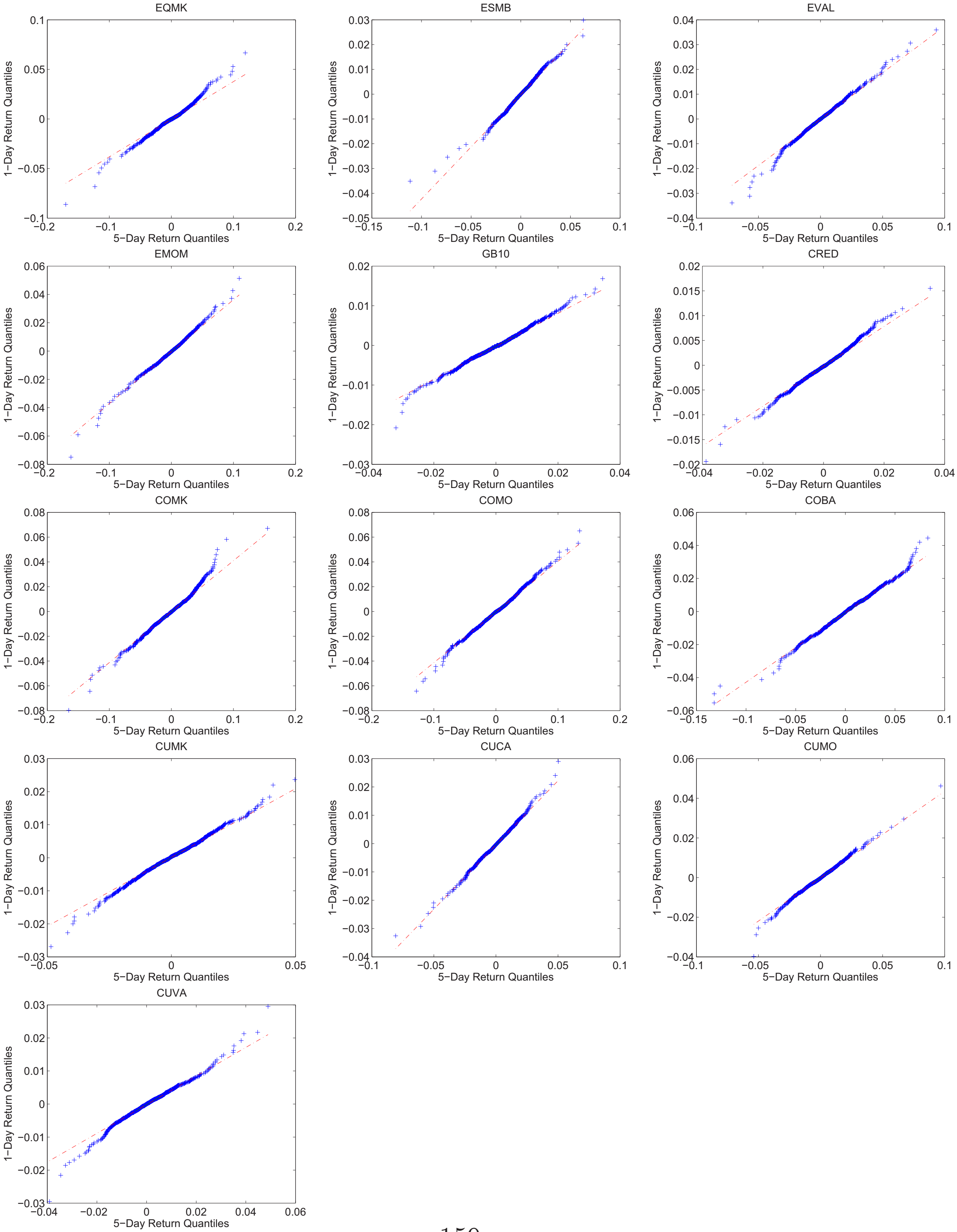
Table 2.19: Statistics of Indices (5-Day Returns)

The table shows the mean (Mean), standard deviation (Std), skewness (Skew) and kurtosis (Kurt) of the index series used in this study. All metrics are calculated by using 5-day log-returns for the period from 31/5/1991 to 26/12/2012.

	EQMK	ESMB	EVAL	EMOM	GB10	CRED	COMK	COMO	COBA	CUMK	CUCA	CUMO	CUVA
Mean	0.105%	0.025%	0.073%	0.120%	0.092%	0.082%	0.002%	0.212%	0.263%	−0.036%	0.110%	0.063%	0.098%
Std	2.521%	1.338%	1.448%	2.347%	0.880%	0.748%	2.993%	3.183%	2.538%	1.208%	1.172%	1.347%	0.947%
Skew	−0.68	−0.81	0.51	−1.24	−0.22	−0.39	−0.54	−0.07	−0.27	0.15	−0.78	0.20	0.41
Kurt	7.59	10.98	7.82	11.03	4.02	4.94	5.51	4.16	5.02	3.89	7.68	6.93	5.39

Figure 2.11: QQ Plots 5-Day Versus 1-Day Returns

The following graphs show the quantile-quantile (QQ) plots of the 5-day return distribution versus the 1-day return distribution for the 13 indices under consideration. The line in each graph connects the first and third quartile of each distribution and is extrapolated to the ends of the distributions. Deviations from the extrapolated line can be interpreted as heavier or lighter tails of one distribution compared to the other.



Chapter 3

Portfolio Insurance Strategies in a Pension Fund Framework

3.1 Introduction

This study examines the question whether common portfolio insurance strategies such as constant proportion portfolio insurance (CPPI) strategies, stop-loss strategies or synthetic put replication strategies are beneficial in the context of a pension fund setting. In this paper a pension fund setting similar to the pension system in Switzerland is being considered. The goal of this study is to examine the utility and risk implications of applying different passive strategies and dynamic insurance strategies in the proposed setting within a simulation framework. Thereby, the system's particularity that the potential restructuring costs of a pension fund may be born by the individual investors (employees) as well as the employer are taken into account. As a particularity the employee's preferences are modeled by expected utility theory functions as well as prospect theory value functions. As it will be shown, depending on the structural parameters of the pension fund and the mapping of the preferences, certain dynamic approaches are dominating other investment approaches with regards to the preferences of the individual fund investor (employee) as well as the employer. This study is inspired by the results of Dichtl and Drobetz (2011) who show that most portfolio insurance strategies are the preferred investment strategy for prospect theory investors when compared to a buy and hold investment strategy. Also, the study is directly linked to the research of Dierkes, Erner and Zeisberger (2010), who find that the attractiveness of portfolio insurance strategies for prospect theory investors depends substantially on the choice of the investment horizon. The results of this study are of practical interest as they justify or deny the use of dynamic insurance strategies versus buy-and-hold strategies in a pension fund setting. Further, the conditions under which a dynamic strategy may be preferred over a passive strategy are retrieved for the given framework.

Related Literature

This study directly extends the work of Dichtl and Drobetz (2011) and the related work on prospect theory developed in Tversky and Kahneman (1992) as well as in Lattimore, Baker and Witte (1992). Further, it is directly related to the work on portfolio insurance strategies, including for example the introduction of the CPPI strategy in Black and Jones (1987). However, it is also related to the string of literature analyzing dynamic asset allocation problems. Related studies in this field are for example: Vigna and Haberman (2001) who solve the asset allocation problem of a discrete time defined contribution (DC) pension scheme for individual plan investors applying dynamic programming techniques to find the optimal investment strategy . Due to the dynamic programming approach, the number of constraints which may be considered in such a setting are limited. A simulation based study considering a similar topic and introducing fairly complex constraints is Sbaraglia et al. (2003). Using a static stochastic optimization approach they are able to solve a complex dynamic asset allocation problem. However, the approach is only feasible for limited investment horizons due to the exponential growth of the optimization problem. More recent research related to this study is Detemple and Rindisbacher (2008), who solve a dynamic asset allocation problem where a parameter for the tolerance of a shortfall in the coverage ratio is being introduced. Also Di Giacinto, Federico and Gozzi (2011), who model a DC pension fund with a minimum guarantee and a solvency constraint are considering a similar problem as the one tackled in this study. While this study has a direct connection to the dynamic asset allocation literature there is a key distinction which needs to be addressed. The study does not try to find the optimal asset allocation strategy of all possible investment strategies as it is often the goal in the dynamic asset allocation literature. The goal is to compare buy and hold strategies with portfolio insurance strategies in the context of a pension fund setting, where the preferences of employees and employers are considered simultaneously and the preferences of the employees may be of the prospect theory type. One advantage of the approaches used in the classical dynamic asset allocation literature is the ability to deal with inter-temporal hedging behavior when there is some predictability in the risk and/or return characteristics of the investment universe. On the other side, the simulation based approach taken in this study has the advantage of being able to model discrete, non-linear characteristics of a typical pension fund setting which are very difficult or impossible to introduce in a classical dynamic asset allocation problem.

The study proceeds as follows: First, in section 3.2.1, the Swiss pension system is explained in more detail. This introduction into the Swiss pension system is followed by a formal description of the proposed framework as it will be used in the simulation part of the study. In section 3.2.2, the utility measurement from the employee's view as well as the employer's perspective are discussed. Section 3.3 continues by explaining the different passive, semi-passive and dynamic strategies which are compared to each other in the context of the given pension fund model. In section 3.4, the simulation setup used in this study is exposed and the simulation results are discussed. Section 3.5 concludes.

3.2 Preference Measurement in the Pension Fund Model

3.2.1 Pension Fund Model

The Funded Private Pension System in Switzerland

The Swiss Pension system consists of three pillars which are described by Queisser and Vittas (2000) as an unfunded redistributive public pillar, a funded occupational pillar and a pillar based on voluntary personal savings. The unfunded public pillar (first pillar) referred to as Alters- und Hinterlassenenversicherung/Assurance Vieillesse et Survivants (AHV/AVS) is designed as a pay-as-you-go system, which is funded by salary based contributions from employees and employers, as well as direct government subsidies and a contribution from the value-added-tax (VAT). The benefits for the retirees from the AHV/AVS have a floor and a cap which correspond to approximately 20 to 40 percent respectively of the average wage in Switzerland (Schulze et al. (2007)). The funded occupational pillar, often referred to as the second pillar, was deemed compulsory in 1985 for all employees whose annual salary exceeded a minimum level. The goal of the second pillar was to provide additional partial coverage for the wage share exceeding 40 percent of the average salary up to 120 percent of the average salary, with the goal of achieving a combined 60 to 70 percent wage replacement rate at retirement (Schulze et al. (2007)). The second pillar is funded by equal contributions from employers and employees, whereby employers may choose to fund more than 50 percent of the contributions due. Unlike the unfunded public pillar, which is centrally organized, the second pillar is provided by individual pension funds which must satisfy minimum legal requirements but are otherwise free to provide more generous benefits and conditions[1]. A noteworthy characteristic of the second pillar is the fact that the employer chooses the

[1] According to the Swiss pension fund study conducted by Swisscanto, a Swiss Asset Manager, there existed 2265 active pension funds in the year 2012 in Switzerland. The study is available under:
http://www.swisscanto.ch/ch/de/berufliche-vorsorge/publikationen/pkstudie/studien.html

pension fund for all its employees, thus leaving individual employees no discretion in choosing another, potentially better performing pension fund[2]. The third, voluntary personal savings pillar, is based on individual contributions made to dedicated retirement accounts which are fully tax deferred up to a specified annual cap[3]. According to Schulze et al. (2007), although the third pillar currently plays a relatively small role in the Swiss pension system its impact is fast-growing.

The focus of this study is on the funded occupational pillar (second pillar), and in order to have a clear understanding of the simulation setup, some additional feature and problems related to the current setup of the second pillar need to be addressed in more detail. As it was mentioned above, the pension funds in Switzerland are subject to certain minimum legal requirements, for example with respect to the minimum interest rate paid on the managed funds as well as the contributions made by the employer. In this study, the focus will be on a setting where the minimum requirements are in place and structures with more generous settings are not further investigated. In general, in the minimum requirement setting, the pension fund is funded by equal monthly contributions from the employer and the employees. Thereby, the contributions are first paid at the age of 25 until the retirement age of 65, whereby the contributions increase over time[4]. Recently, the increasing contribution over time have been criticized by politicians, as this may lead to discriminating hiring practices as younger people are preferred due to the lower ancillary wage costs[5]. The contributions made every month, are credited to an individual account for each employee. On the funds ac-

[2]However, the pension fund needs to be a separate legal entity, in order to avoid funding gaps should the employer default.

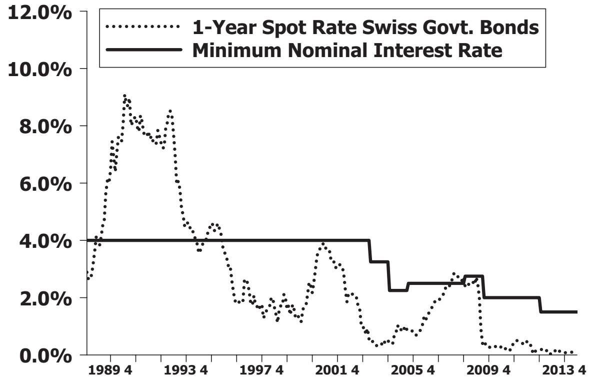
[3]The maximum annual deferred amount for individuals covered by an occupational pension is 6'739 CHF in 2013. For individuals not affiliated with a pension fund, for example self-employed workers, the maximum deferred amount is set to 20% of the net earned income with a cap at 33'696 CHF in 2013.

[4]Currently, the following contribution in percentage of the covered wage are made: Age 25-34: 7%, Age 35-44: 10%, Age 45-54: 15%, Age 55-65: 18%. Retirement before the age of 65 is possible at the cost of reduced benefits.

[5]See for example:
http://www.news.admin.ch/NSBSubscriber/message/attachments/3824.pdf

Figure 3.1: Minimum Nominal Interest Rate versus Actual 1-Year Swiss Government Bond Yield

Minimum nominal interest rate (Mindestzins) as set by the Swiss Federal Council as well as the 1-Year Swiss Government Bond Spot Yield for the time period of January 1988 until December 2013.



cumulated by the employee, a minimum nominal interest rate needs to be paid by the pension fund. The minimum nominal interest rate is set on a discretionary basis by the Swiss Federal Council (Bundesrat) and is being reviewed at least bi-annually. While the law stipulates that the financial market environment needs to be accounted for when setting the rate, the recent history has shown that the tie to the current market environment is loose and that the rate-setting is heavily influenced by political motivations. Figure 3.1 shows the minimum nominal interest rate as well as the 1-year Swiss government bond yield for the period of January 1988 until December 2013.[6] Given the minimum nominal interest rate, a projection can be made for the accumulated capital by age 65. This projection is provided to every employee on an annual basis and this projection will be key to determine the reference point needed when determining the benefit of an employee when the preferences are mapped by prospect theory. The accumulated capital is converted into a life annuity at the age of 65, whereby the retiree has the option to retrieve the capital instead of converting it into an annuity. The terms of the annuity are determined by the government as well, whereby the

[6] Sources: Homepage Swiss Confederation (http://www.admin.ch/opc/de/classified-compilation/ 19840067/index.html#a12) and homepage Swiss National Bank (http://www.snb.ch/ de/iabout/stat/statpub/id/statpub) respectively.

so called conversion ratio is a key element. The conversion ratio expresses the annual pension benefit in percentage points of the accumulated capital which will be paid out to the retiree. From the year 2014 onwards, the conversion ratio is set at 6.8%. In the current system, there is only a very limited annuity market as the conversion ratio is set on a discretionary basis by the government and the pension funds carry the full risk of the annuity liabilities.

A final consideration which needs to be addressed are the procedures which are initiated once a pension fund is under-funded, meaning that the coverage ratio, defined as assets divided by liabilities, is smaller than one. While the law states that when a pension fund is under-funded, it must take action to re-install full coverage, it is possible for pension funds to have a coverage ratio below 100% without having to reduce the benefits of the insured employees. The precondition for this is a credible plan how the fund will regain a coverage ratio above 100% in the foreseeable future, for example by adjusting the investment strategy. However, should the coverage ratio fall below 90%, this is often seen as a considerable under-funding and restructuring measures need to be installed[7]. These measures may include lump sum payments by the employee and the employer as well as a reduction of the provided interest rate to a level below the minimum nominal interest rate for a predefined period of time. In the case of lump sum payments, the employer needs to contribute at least 50% of the total amount due. With respect to the minimum coverage ratio, the pension funds of government-related entities may opt for a partial-capitalized model where the minimum coverage ratio is set at 80%. Thus, depending on the pension fund setting, minimum coverage ratios above 80% to 90% need to be upheld in order to avoid benefit reducing restructuring measures.

[7] See for example the resources on the homepage of Weibel Hess & Partner, as pension fund consultancy firm, for a more detailed discussion of the issue: http://www. pensionskassenvergleich.ch/ pkvergleich/pkvergleich2009/sanierungsmassnahmen/index.php

Applied Pension Framework

The proposed framework is very similar to the current system in place, whereby two element differ marginally from the actual system. The two adjustments to the system are as follows:

The first adjustment is related to the setting of the minimum nominal interest rate. Instead of relying on a discretionary fixed and updated minimum interest rate, a low fixed rate r_{fix} is assumed which will not be altered over the course of time. This rate could be chosen according to the long term inflation expectation or the central banks inflation target of 2%. Then, depending on the performance of the pension fund investment strategy, extra dividends or restructuring outlays will be added or subtracted to this minimum nominal interest rate. In particular, for the simulation conducted in this study, an extra dividend is added to the fixed minimum rate whenever the coverage ratio of the pension fund exceeds a coverage ratio of $CR_{max} = 120\%$ by year end. The extra dividend will be chosen such that after the payout of the extra dividend, the coverage ratio of the fund will equal CR_{max} again. In the case that the coverage ratio falls below a minimum coverage ratio CR_{min}, restructuring outlays will be subtracted from r_{fix}, such that a new post restructure coverage ratio $CR_{reset} = 95\%$ is obtained. Thereby, only half of the total restructuring outlays are debited from the employees accounts, while the rest is contributed by the employer which corresponds to the current regulation where the employer has to carry at least 50 percent of the restructuring costs if lump sum payments are used.

The second adjustment is made with respect to the treatment of the annuity liabilities. In the proposed system the capital pool of the employees and the retirees is being separated. In particular, it is assumed that the retirees are paid-out the accumulated capital at the point of retirement and decide whether they want to convert it into a life-annuity in the private market at the present market rate. This assumption significantly simplifies the simulation setup as only the time until retirement needs to be simulated. Further, certain simplifying assumptions are made in order to keep the simulation framework lean. One such assumption is that the contributions made by employees are constant over time, and are not increasing with their age as it is currently the case. Further, instead of monthly contributions, annual contributions are assumed in the simulation framework.

Current Pension Reform Discussion

During the past decade, the need for reforms to the second pillar in the Swiss pension system has become more pressing. Two of the main problems of the current system are the discretionary determination of the minimum nominal interest rate as well as the conversion ratio for the life annuity. Among pension fund experts in Switzerland there is currently a consensus, that the current conversion ratio of 6.8% is too high due to the current interest rate environment and the higher than anticipated life expectancy. The too high conversion ratio leads to a wealth redistribution from the active employed generation to the retired generation. Also, the minimum nominal interest rate has not been able to properly reflect the market environment and has been adjusted to the changing environment only with a significant lag. As a consequence, the calls for a reform have become louder and recently for the first time an approach has been initiated in the Swiss lower house of parliament to develop a framework to adjust the conversion ratio as well as the minimum nominal interest rate in an automated rule-based fashion[8]. However, the path to an actual change in the law is long and will likely have to pass a popular vote later in the process. The next section gives a formal description of the pension framework used in this study.

[8] The approach was submitted in the form of a motion directed at the Swiss Federal Council on September 11th 2013 to work out a proposal for a rule based approach for setting the conversion ratio as well as the minimum nominal interest rate.

Formal description of the pension framework

In order to formally describe the proposed pension fund system used in this study I start by describing how the accounts of the investors (employees) are modeled. Instead of simulating individual investor accounts, the accounts of every age cohort are aggregated. Thus, instead of simulating the individual investor accounts of all investors retiring in a specific year, the accounts are aggregated and represented by a single cohort investor account. In effect, there are $N = 40$ cohort investor accounts active at any point in time, representing the individual investors aged 25 to 64, with a time to retirement between 40 and 1 years. The cohort investor accounts are denoted by the scalar y, which indicates the year when the cohort retires. For example, the cohort $y = 5$, will retire in 5 years time. The variable t denotes the year indicator. As will be explained in more detail in section 3.2.2, the utility respectively the prospect theory value (PV) of the cohort investor y, realized in simulation path m, depends on the difference between the actual payout and the expected promised payout $\Delta x_{y,m}$. Thereby, $\Delta x_{y,m}$ is defined as follows, where $IA_{y,m,t}$ is the actual account value of the cohort investor y at the time of maturity $t = y$ and $EA_{y,m}$ is the expected account value:

$$\Delta x_{y,m} = IA_{y,m,t}/EA_{y,m} - 1 \quad where \quad t = y \tag{3.1}$$

In order to simplify the notation, the scenario subscript m will be dropped in the following formal description. The expected account value EA_y is defined as the sum of the annual contributions c over the investment horizon N, compounded at the fixed rate r_{fix}. The annual contribution c of every cohort investor are chosen to equal 25 monetary units[9]. The actual investor account value $IA_{y,t}$ is determined by the actual compounded rates r, which may vary year over year, and the initial actual account value $iniIA_y$, which represents accrued contributions paid before the first observed year 0 (see equation 3.3).

[9]In this setting, the expected account value equals $40 \cdot 25 = 1000$ monetary units when $r_{fix}$is set equal to 0.

It is assumed that the contributions made in the initial actual account $iniIA_y$ have been accrued by the fixed rate r_{fix} (see equation 3.4)[10]. The variable d_y, is a weighting factor which accounts for demographic differences between the different cohort investors (e.g d_y is higher when the specific cohort is over-represented).

$$EA_y = \sum_{u=1}^{N} d_y c \prod_{s=u}^{N} (1 + r_{fix}) \tag{3.2}$$

$$IA_{y,t} = iniIA_y \prod_{j=1}^{t} (1 + r_j) + \sum_{j=u}^{t} d_y c \prod_{s=j}^{t} (1 + r_s) \tag{3.3}$$

$$where\ u = t - min\,(y, N) + 1$$

$$iniIA_y = \begin{cases} \sum_{j=1}^{N-y} d_y c \prod_{i=j}^{N-y} (1 + r_{fix}) & if\ 1 < y \leq N \\ 0 & otherwise \end{cases} \tag{3.4}$$

The actual return r is dependent on the coverage ratio of the pension fund at the end of the year t. The coverage ratio of the pension fund at the end of year t is calculated by dividing the fund assets FA_t by the fund liabilities FL_t whereby it is assumed that the fund liabilities are compounded by the fixed rate r_{fix} for year t. The definition of FL_t is given in the following equation:

$$FL_t = \sum_{y=t}^{t+N-1} IA_{y,t}^{pre} \tag{3.5}$$

$$IA_{y,t}^{pre} = (iniIA_y \prod_{j=1}^{t-1} (1 + r_j) + \sum_{j=u}^{t-1} d_y c \prod_{s=j}^{t-1} (1 + r_s) + d_y c)(1 + r_{fix}) \tag{3.6}$$

$$where\ u = t - min\,(y, N) + 1$$

[10]For example, the cohort investor $y = 5$, has an initial actual account value of 875 monetary units when r_{fix} is set at 0%, which corresponds to 35 years of annual contributions c.

Whenever the coverage ratio of the fund at the end of the year $CR_{end,t}$ lies between the minimum coverage ratio CR_{min} and the maximum coverage ratio CR_{max}, the account value is compounded at the fixed rate r_{fix}. Otherwise, the account value is adjusted such that the reset coverage ratio CR_{reset} respectively the maximum coverage ratio CR_{max} is obtained. When the maximum coverage ratio is exceeded, this is done by distributing an extra dividend $r_{extra,t}$ to the investor accounts, such that with the increased liabilities the coverage ratio of the fund is again set to CR_{max} (third line equation 3.7). In the case of an under-funding ($CR_{end,t} < CR_{min}$), the reset level is obtained by reducing the fixed rate by a restructuring contribution $r_{restruc,t}$. Thereby, $r_{restruc,t}$ is determined by the fund assets FA_t, the fund liabilities FL_t, the reset coverage ratio CR_{reset} as well as the restructuring share RS, which determines the contribution which needs to be made by the employee. For this study, RS is chosen to equal 50%. As the employee only contributes a partial amount needed to recover to the reset coverage ratio by reducing the liabilities, the employer also needs to increase the assets by a corresponding lump-sum payment. The lump-sum payment made by the employer is dependent on the fund liabilities after accounting for employee restructuring contributions, denoted FL_t^{end} (equation 3.8), the fund assets FA_t, as well as the reset coverage ratio CR_{reset}. The employer contribution for year t denoted EC_t is defined in equation 3.9.

$$r_t = \begin{cases} r_{fix} & if\, CR_{end,t} \geq CR_{min} \quad and \quad CR_{end,t} \leq CR_{max} \\ r_{fix} + r_{restruc,t} & if\, CR_{end,t} < CR_{min} \quad where \quad r_{restruc} = RS \cdot \left[\left(\frac{FA_t}{CR_{reset}} \right) / FL_t - 1 \right] \\ r_{fix} + r_{extra,t} & if\, CR_{end,t} > CR_{max} \quad where \quad r_{extra} = \left[\left(\frac{FA_t}{CR_{max}} \right) / FL_t - 1 \right] \end{cases} \tag{3.7}$$

$$FL_t^{end} = \sum_{y=t}^{t+N-1} IA_{y,t} \tag{3.8}$$

$$EC_t = \begin{cases} 0 & if\ CR_{end,t} \geq CR_{min} \\ CR_{reset} \cdot FL_t^{end} - FA_t & if\ CR_{end,t} < CR_{min} \end{cases} \quad (3.9)$$

The fund assets FA_t as described in equation 3.10 correspond to the fund assets at the year end before any adjustments are made. They are calculated iteratively by taking the fund assets at the end of the prior year FA_{t-1}^{adj}, adding the annual contributions, which are made at the beginning of the year, and compounding the total assets by the realized rate of return $r_{realized,t}$[11]. FA_{t-1}^{adj} as defined in equation 3.11, is calculated by adjusting FA_{t-1} for retirement outflows $IA_{t-1,t-1}$ and employer restructuring contributions EC_{t-1}.

$$FA_t = \left(FA_{t-1}^{adj} + Nc\right)(1 + r_{realized,t}) \quad (3.10)$$

$$FA_t^{adj} = FA_t - IA_{t,t} + EC_t \quad (3.11)$$

The initial values for the fund assets FA_0 and the fund liabilities FL_0 are defined as follows, where the fund liabilities are equal to the fund assets (initial coverage ratio of 100% is assumed) and the liabilities are determined by the sum of the individual investors initial account values.

$$FA_0 = FL_0 \quad (3.12)$$

$$FL_0 = \sum_{y=1}^{N} iniIA_y \quad (3.13)$$

3.2.2 Preference Measurement

The goal of this study consists of investigating whether and when dynamic trading strategies may be preferred over passive approaches in the proposed pension fund setting. Thereby, not only the view of the fund investor/employee is being considered, but also the view of the employer who is liable at least in part for the contributions to a pension fund restructuring

[11] The realized return $r_{realized,t}$ depends on the market development over the course of year t as well as the chosen investment strategy.

in the given framework. With regards to the employee's view, the concepts of prospect theory as well as classical expected utility theory are applied to determine the attractiveness of the considered strategies. On the other side, when analyzing the employer's view, it is assumed that the employer is risk neutral such that a measure of expected restructuring costs is relevant for determining the attractiveness of the different strategies.

Fund Investor View - Prospect Theory and Expected Utility Theory

The studies of Dierkes, Erner and Zeisberger (2010) and Dichtl and Drobetz (2011), have shown that for individual investors, dynamic portfolio insurance strategies may be preferred over passive strategies when mapping the preferences of investors by prospect theory instead of classic expected utility theory. In this study I follow the prospect theory approaches applied in Dierkes, Erner and Zeisberger (2010) and Dichtl and Drobetz (2011) and extend them to the proposed pension fund setting where multiple cohort investors with different investment horizons are invested in a single strategy. In order to determine the impact of choosing prospect theory versus classical expected utility theory (EUT), the results are also determined with CRRA utility functions. It will be shown that the choice of the preference framework has a significant impact on the overall results.

Prospect theory tries to map the empirical behavior of individuals when choosing between alternatives that involve risk (Dichtl and Drobetz (2011)). Key empirical findings, for example in Tversky and Kahneman (1992), indicate that individuals evaluate the alternative choices based on expected gains and losses relative to a reference point. In EUT it is assumed that investors only care about the absolute value of the payoff irrespective of a reference point. Further, the empirical findings show that individuals are risk averse in the domain of wealth gains and risk seeking in the domain of wealth losses relative to the reference point. This leads to what is sometimes referred to as an S-shaped value function compared to the strictly concave utility function used in EUT. A further important feature reflected in the prospect theory approach is related to the observed risk aversion of individuals. This means that a loss, compared to a gain of the same magnitude, has a stronger effect on the value function. A common specification to capture these features is provided by the two-part valuation function in Tversky and Kahneman (1992) as shown in equation 3.14.

$$v(\Delta x) = \begin{cases} (\Delta x)^{\alpha} & if\ \Delta x \geq 0 \\ -\lambda(-\Delta x)^{\beta} & if\ \Delta x < 0 \end{cases} \tag{3.14}$$

In the above equation, the parameters α and β determine the curvature of the value function for gains respectively losses and λ defines the risk aversion of the investor. The values chosen for the three parameters are 0.88, 0.88, and 2.25 respectively. Thereby, the chosen values correspond to the values estimated in Tversky and Kahneman (1992) and are also the values chosen in Dichtl and Drobetz (2011)[12]. The $\triangle x$ variable in the above equation corresponds to the gain respectively loss relative to the chosen reference point. In the studies of Dierkes, Erner and Zeisberger (2010) and Dichtl and Drobetz (2011), the reference point is chosen to equal zero and Δx is defined as the simple return of the asset over the investment period[13]. In the setting of this study, Δx is chosen to be the percentage deviation of the actual payout compared to the expected payout of a specific cohort investor as defined in equation 3.1 in section 3.2.1[14]. Choosing the expected payoff at retirement seems to be a reasonable choice for the reference point, as it is a figure which is communicated to the employees on an annual basis and may be used by individuals to adapt a consumption and savings plan accordingly.

[12] A broad sensitivity analysis for the parameter choice for the value function in the context of choosing among different investment strategies can be found in Dierkes, Erner and Zeisberger (2010).

[13] Their approach assumes that the reference point for the investor is the purchase price of the security.

[14] Choosing absolute deviations instead of percentage deviations does not qualitatively alter the results presented in this study.

In order to determine the aggregate value of a certain strategy for all employees, the prospect values of all employees are summed up as described in the following equation:

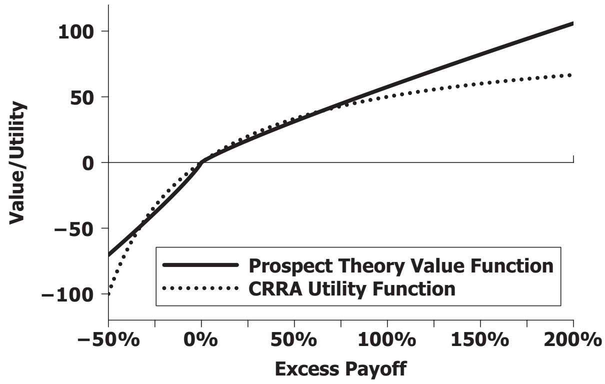
$$AggPV(Strategy) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{y=1}^{N} d_y \cdot v(\Delta x_{y,m}) \tag{3.15}$$

Thereby, $N = 40$ which means that only the utilities of the employees who are in the fund at the start of the simulation are being considered in the value aggregation. The variable d_y, is again the weighting factor which accounts for demographic differences between the different cohort investors. Using equation 3.15, the aggregate utility of the different strategies under consideration which are described in section 3.3, can be compared to each other on the assumption that the utilities of the cohort investors are described by a prospect theory value function. A further empirical finding of prospect theory states that individuals are not using statistical probabilities to weight different outcomes, but show a tendency to overweight extreme outcomes (low probability events) over normal outcomes (high probability events). In order to evaluate the effect of subjective probability weighting, Dichtl and Drobetz (2011) for example apply a probability weighting function as proposed in Lattimore, Baker and Witte (1992). However, since the estimation of the parameters for the probability weighting function is controversial and problematic as laid out for example in Tversky and Kahneman (1992), and in order to reduce the complexity of the problem at hand, in this paper results using cumulative prospect theory are not shown[15].

Besides using the prospect theory value function, also an alternative specification based on EUT is used, whereby a constant relative risk aversion (CRRA) utility function is used as described in equation 3.16. Thereby, the percentage deviation Δx is shifted such that it represents the payout level

[15] Results for this study using the parameter specifications found in Abdellaoui (2000) indicate that introducing a probability weighting function exaggerate the differences when comparing the outcomes between prospect theory and expected utility theory specifications.

Figure 3.2: Prospect Theory Value Function and CRRA Utility Function
This figure shows the prospect theory value function as well as the CRRA utility function as defined in equations 3.14 and 3.16 respectively. The parameter are set as follows: $\alpha = \beta = 0.88$, $\lambda = 2.25$ and $\kappa = 2$.



in percentage terms of the expected payoff. Further, the utility is shifted by adding the summand $1/(\kappa - 1)$ which ensures that the utility is equal to 0 whenever the actual payoff is equal to the expected payoff[16]. Also, the results are scaled by a factor of 100. The risk aversion parameter κ is chosen to equal 2 for the purpose of this study. For the calculation of the aggregated utility, the same approach as in equation 3.15 is used, whereby $v(\Delta x_{y,m})$ is replaced with $CRRA(\Delta x_{y,m})$. In figure 3.2, plots of the prospect theory value function as well as the scaled CRRA utility function are shown given the excess payout Δx. The parameter choices for the functions shown in the figure correspond to the ones used in this study defined above.

$$CRRA(\Delta x_{y,m}) = \left(\frac{(\Delta x_{y,m} + 1)^{1-\kappa}}{1 - \kappa} + \frac{1}{\kappa - 1} \right) \cdot 100 \tag{3.16}$$

Employers View - Expected Restructuring Costs

As it was explained in section 3.2.1, in the proposed pension fund system the employer is held liable for restructuring costs whenever the coverage ratio of the pension fund falls below a predefined minimum coverage ratio CR_{min} by year end. For the purpose of this study it is assumed that the

[16]Shifting the utility has no impact on the overall results and using absolute values instead of Δx does not alter the results in a qualitative way.

employer absorbs 50% of the restructuring costs. Further, it is assumed that the employer is risk neutral with regards to the expected restructuring contributions ERC. Therefore, the utility from the perspective of the employer can simply be determined by calculating the mean expected restructuring costs over the horizon of $N = 40$ years for M simulation runs as done in the equation below. Thereby, the employer contribution EC_t is defined as in equation 3.9 and r_{long} corresponds to the long term expected average interest rate and is used to discount the future liabilities[17].

$$ERC(Strategy) = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=1}^{N} (1 + r_{long})^{-t} (-EC_t) \tag{3.17}$$

Similar to the $AggPV$ as well as the $CRRA$ specifications for the employees perspective, the expected restructuring contribution (ERC) as defined in equation 3.17 allows to determine the preference for specific strategy from the viewpoint of the employer. As an alternative, the employer may be interested not in the mean expected restructuring contributions but for example the costs which are exceeded in 5% of the worst cases. This approach is a value-at-risk (VaR) approach as it is commonly applied in risk management and is described in the equation below:

$$ERCPerc(Strategy) = Q_p \left(\sum_{t=1}^{N} (1 + r_{long})^{-t} (-EC_t) \mid \mathbf{M} \right) \tag{3.18}$$

In the above equation Q_p is the quantile function covering the set of all possible outcomes $\mathbf{M}$, and p defines the threshold below which p% of all observations fall. If the employer is more concerned with regards to the tail risk of possible future restructuring contributions, the expected restructuring cost percentile approach ERCPerc, as described in equation 3.18, may be more suitable to reflect the utility from the employer's point of view. In this

[17] The interest rate r_{long} is set equal to the average risk free rate corresponding to the various scenarios outlined in section 3.4.

study p is set at the 5% level.

3.3 Considered Investment Strategies

The strategies considered in this study can be divided into passive, semi-dynamic and dynamic strategies. Thereby, the dynamic strategies considered in this study are forecast-free insurance strategies where no expectations about the development of future security prices are needed to determine the allocation between the risk free asset (cash) and the risky asset (equity). All considered dynamic insurance strategies have the goal to protect a certain level of wealth over a given period of time and are similar to those analyzed for example in Annaert, Osselaer and Verstraete (2009), Dierkes, Erner and Zeisberger (2010) and Dichtl and Drobetz (2011).

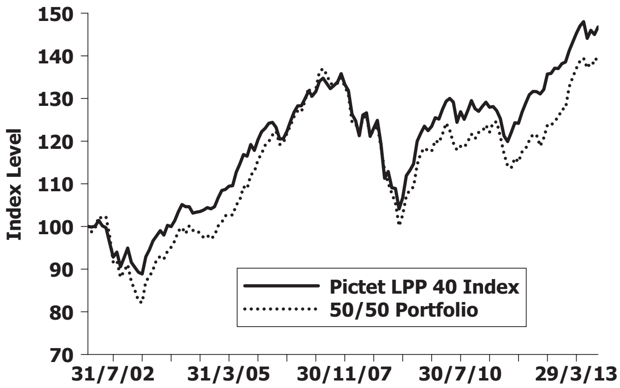
3.3.1 Passive and Semi-Dynamic Strategies

Two passive and one semi-dynamic strategy are being considered in this study. The two passive strategies are a full investment in equity (FEQ) and a full investment in cash (CASH). These two strategies do not require any re-balancing over time and the returns are simply the return of the equity market, respectively those of a short term cash deposit. The semi-dynamic strategy considered in this study is a 50/50 allocation between the equity asset and the cash deposit (50/50). Whereas it is possible to fix this allocation at the beginning of the observation period and hold the assets without re-balancing, in this study I choose to rebalanced the portfolio at the end of each year. Thus, at the end of every calendar year the allocation is being rebalanced such that a 50/50 allocation is again retrieved. The semi-dynamic 50/50 allocation can be regarded as a proxy for the actual strategy being deployed by many pension funds[18]. The comparison of the

[18] The 50/50 allocation simplifies the actual strategy deployed by many pension funds in particular with regards to two aspects. First, no term premium is being considered, which is a compensation for holding longer duration bonds. Second, the credit spread earned for example

Figure 3.3: Pictet 40 Index versus 50/50 Allocation

Figure showing the Pictet LPP 40 Index versus a 50/50 allocation semi-dynamic strategy which allocates 50% of the assets to the Swiss Performance Index (SPI) and 50% to the Pictet Short Term Money Market Fund with monthly rebalancing for the time period January 2001 until September 2013.



Pictet 40 pension fund index and a balanced approach which allocates 50% of the assets to the Swiss Performance Index (SPI) and 50% to a Swiss Franc Money Market Fund with monthly re-balancing is shown in figure 3.3[19].

3.3.2 Dynamic Insurance Strategies

The three dynamic insurance strategies considered in this study are a constant proportion portfolio insurance strategy (CPPI), a put option replication strategy (PORS) and a stop-loss strategy (SLS). All three strategies have in common that an asset floor level or strike level below which the portfolio value should not fall needs to be determined. In the context of the chosen pension fund setting it is an obvious choice to link the strike level SL_t to the minimum coverage ratio CR_{min} as well as the promised fixed rate r_{fix}. In particular, the strike level is reset at the start of every year and chosen such that by year end, before any distributions are made, the pension fund coverage ratio is above or equal to CR_{min}. In the equation below, the variable $CR_{start,t}$ refers to the coverage ratio at the beginning of the year t

on corporate bonds is being neglected. However, with regards to the credit spreads it may be argued that the underlying risk is very similar to the equity market risk, and that therefore the risk and return of a pure credit portfolio can be proxied by the latter.

[19] Data from Bloomberg, using the following tickers: LPPD40 Index, SPI Index and PIPCHRI LX Equity.

and the maximum strike level is capped at 100%[20].

$$SL_t = min\left(\frac{CR_{min}(1 + r_{fix})}{CR_{start,t}}, 100\%\right) \tag{3.19}$$

The three considered strategies have been analyzed in various previous studies with respect to their payoff properties[21]. The least sophisticated strategy is the SLS, which is fully invested in either the risky asset or the risk free asset at all times. PORS is more sophisticated as the strategy reduces the exposure to the risky asset in a continuous fashion as the strike level is approached. However, falling below the set floor is common with this strategy. The CPPI strategy also continuously reduces the exposure to the risky asset as the floor level is approached, but the probability of actually falling below the desired floor is very low. Thus, the three strategies all reduce the exposure to the risky asset when the portfolio value decreases but they do it in distinguished ways which lead to different payout structures. The detailed workings of the three strategies are explained in the next subsections.

Constant Proportion Portfolio Insurance Strategy (CPPI)

The CPPI strategy was introduced by Black and Jones (1987) and Black and Jones (1988)[22]. The idea of the strategy consists of protecting the wealth of an investor by determining a certain floor below which the portfolio value should not fall and then allocating the wealth dynamically between the risky asset, equity in our case, and cash. In this study the floor is referred to as the strike level SL_t. The percentage amount of the portfolio value which is allocated to the risky asset when using the CPPI strategy is determined by

[20] Otherwise SL_t may be larger than 100% in the case where the values of CR_{min} and $CR_{start,t}$ are almost equal and $r_{fix}$is high.

[21] See for example Annaert, Osselaer and Verstraete (2009), Dierkes, Erner and Zeisberger (2010) and Dichtl and Drobetz (2011).

[22] One of the early studies investigating the payoff properties of the CPPI strategy is Black and Perold (1992), who show that it is the optimal investment strategy when a piecewise-HARA utility function with a minimum consumption constraint is assumed.

the cushion C_t. The cushion is defined as follows:

$$C_t = PV_t - SL_t \tag{3.20}$$

Thereby, PV_t represents the portfolio value at time t, as a percentage of the initial portfolio value at the beginning of the year. The weight exposure w_t to the risky asset at each point in time is determined by multiplying the cushion C_t by a multiplier mu. The remainder of the wealth not invested in the risky asset is held in cash. In order to avoid a leveraged outcome when the cushion becomes very large, the maximum weight is capped at 100%.

$$w_t = min\,(C_t mu, 100\%) \tag{3.21}$$

The multiplier mu is chosen to equal 5 in this study, the same as in Dichtl and Drobetz (2011)[23]. In order to gain a better understanding of mu, it is useful to look at the inverse $1/mu$ which allows for a straight forward interpretation. The inverse $1/mu$ represents the maximum drop which may occur between re-balancing dates, such that the strike level is not breached. Thus, when choosing $m = 5$, the strike level SL_t will not be undercut as long as the risky asset does not drop by more than $1/m = 20\%$. When it is assumed that the market is trading in a continuous fashion, showing no jumps, and that the weight w_t is continuously adjusted, the strike level will never be breached. In order to account for a more realistic setting in this study, rebalancings are only possible at the daily market close with a lag of one day. This means that the adjustment needed to obtain the weight exposure as defined in equation 3.21 for time t is implemented in the close of day $t + 1$. Further, in order to prevent daily trading as a result of minor changes in w_t, trading filters are implemented similar to previous studies. Instead of choosing trading filter thresholds based on market movements in

[23] A recent study of Zieling, Mahayni and Balder (2014) finds evidence that using a variable multiplier may significantly improve the performance relative to a classical CPPI approach. Including variable proportion portfolio insurance strategies (VPPI) in a pension fund setting may be an interesting field for further research.

order to adjust w_t, as done for example in Dichtl and Drobetz (2011), in this study the implied levels of mu are used as trading thresholds. Thereby, the implied levels of mu are found by inverting equation 3.21 at the end of each trading day as follows: $mu_{imp,t} = \frac{w_t}{C_t}$. If $mu_{imp,t}$ is larger or smaller than $mu_{max} = 5.5$ respectively $mu_{min} = 4.5$, a rebalancing is implemented at the close of the next day. Trading costs are accounted for in each trade by charging a round-trip fee of 0.05% of the traded notional.

Put Option Replication Strategy (PORS)

The Put Option Replication Strategy (PORS), is based on the synthetic put strategy proposed in Rubinstein and Leland (1981). The approach replicates the payoff of a portfolio consisting of a risky asset S and a put option on the risky asset $P(K)$ where the strike level K is chosen, such that the desired floor strike level SL_t is protected. In order for SL_t to be protected, the strike level of the put option K needs to be determined by iteratively searching for the strike level K, such that the following relationship holds:

$$K = SL_t(S + P(K)) \tag{3.22}$$

In effect, a strike level K is determined such that accounting for the cost of the option $P(K)$, the floor SL_t is never breached if it is assumed that the option $P(K)$ is an actual option and not a replicated option. In this study however, $P(K)$ represents a synthetic option which is replicated at discrete points in time and there exists the risk of falling below the floor strike level SL_t. The value of the portfolio consisting of the risky asset and the put option is determined by applying the Black and Scholes (1973) option pricing framework as shown in Dichtl and Drobetz (2011).

$$S + P(K) = SN(d1) + Ke^{-rT}N(-d2) \tag{3.23}$$

$$where$$

$$d1 = \frac{ln\,(S/K) + (r + 0.5\sigma^2)T}{\sigma\sqrt{T}} \; and \; d2 = d1 - \sigma\sqrt{T}$$

In the above equations, $N(\cdot)$ is the standard normal cumulative distribution function, r is the annualized risk free rate, T is the time to maturity measured in years, and σ is the volatility of the risky asset. The choices for r and σ are discussed in section 3.4. The delta of the portfolio $S + P(K)$, which is needed for synthetically replicating the payoff of the portfolio, is given in the equation below:

$$\frac{\partial\,(S + P(K))}{\partial S} = N(d1) \tag{3.24}$$

As shown in Dichtl and Drobetz (2011), multiplying the delta given in equation 3.24 by the price of the risky asset S and dividing the result by the value of the total portfolio given in equation 3.23, the percentage weight invested in the risky asset is retrieved. Thus, the optimal weight invested in the risky asset for the PORS at each point in time is defined as follows:

$$w_t = \frac{S_t N\,(d1)}{S_t + P(K)_t} \tag{3.25}$$

Similar to the case of the CPPI strategy, one day implementation lags, the same trading costs and trading thresholds are introduced for the implementation of PORS. In particular, the weight invested in the risky asset is only adjusted when the absolute difference between the optimal weight determined in each point in time by equation 3.25 and the actual weight is larger than 2%.

Stop-Loss Strategy (SLS)

The SLS is the simplest approach to protect the portfolio from falling short of the determined strike level SL_t. The portfolio value PV_t at the beginning of the year is set equal to 100% for reference purposes and the strategy is fully invested in the risky asset, which is an equity index in our case, unless the PV_t falls below SL_t at any time during the year. Should PV_t fall below SL_t at the end of any day, the entire assets are shifted from the risky asset to the cash deposit. In order to have a conservative and realistic setting, the shift from the risky asset to the cash deposit is conducted with a 1-day lag. This means that when a breach of the floor is observed, the shift into cash is conducted in the market closing of the next day. In the study of Dichtl and Drobetz (2011), the SLS is implemented such that the portfolio is invested in cash until the end of the year. Thus, once the strike level is reached, there is no chance to re-enter the market by shifting back to the equity market. In this study, such a shift back into the risky asset is possible, as soon as the portfolio recovers and $PV_t \geqq SL_t$. However, the shift back is again implemented with a one day lag. The same trading costs as for the CPPI strategy are applied.

3.4 Simulation Study Results

Simulation Setup

In order to determine the relative attractiveness of the proposed strategies from an employee's as well as an employer's point of view, a Monte Carlo simulation study is being conducted. Thereby, 40 years of daily returns for two components are simulated: One is the simulation of the equity market returns and the other one is the simulation of the stochastic interest rates[24]. The number of runs M is set to 5000 and it is assumed that the correlation between the two series is zero. The equity market return is simulated by generating continuously compounded stock market returns for each day assuming that the return follows a Geometric Brownian motion defined below where S refers to the stock market price, μ is the drift term, σ_{equity} is the

[24] It is assumed that one year has 252 trading days.

equity market volatility and W_t is a Wiener process.

$$d(lnS_t) = \left(\mu - \frac{\sigma^2_{equity}}{2}\right) dt + \sigma_{equity} dW_t \tag{3.26}$$

Following the assumption in Dichtl and Drobetz (2011), in the base case scenario an expected annual equity market return of 9% is assume[25]. The annualized volatility is set at 17%. In contrast to Dichtl and Drobetz (2011), the interest rate r is not fixed but is assumed to also follow a stochastic process. The process chosen for the interest rate behavior is the CIR-process as defined in Cox, Ingersoll Jr and Ross (1985) and shown in the equation below.

$$dr_t = a(b - r_t)dt + \sigma_{rate}\sqrt{r_t}dW_t \tag{3.27}$$

The mean reversion parameter a, the long term mean value b and the volatility σ_{rate} are set to 0.2, 4% and 10% respectively. Setting the long term mean value of the interest rate equal to 4% may seem relatively high, given the current market environment with record low risk free interest rates near 0% in many developed countries across the world. On the other side, Dichtl and Drobetz (2011) use a rate of 4.5%, referring to the results of Arnott and Bernstein (2002). Looking at the one-month US treasury bill rates for the time period between July 1926 and August 2013, the average rate has been 3.5%, while the data series of Robert Shiller, covering the US market for the period from 1871 to 2011 reveals an annual one-year interest rate of 4.7%[26]. Choosing a long term mean value of 4% therefore represents a middle ground of the values discussed in the literature. Also, the starting level of the interest rate is set to 4%. Alternative interest rate scenarios will be discussed as part of the current environment scenarios in subsection 3.4.3.

[25] This return is transformed into a continuously compounded return in the simulation procedure.

[26] The data for the treasury bill series corresponds to the one used in Fama and French (1993) and is available on Kenneth French homepage:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

The data series of Robert Shiller has been used in Shiller (1992), and is available under the following link: http://www.econ.yale.edu/~shiller/data.htm

Further, in subsection 3.4.4 an alternative setup will be analyzed where a block-bootstrap procedure is applied instead of the Monte Carlo approach.

3.4.1 Base Case Scenario

In the base case scenario the simulation uses the parameters defined above and the parameters for the pension fund corresponds to the ones discussed in section 3.2.1. In particular, the fixed interest rate r_{fix} is set equal to 2%, which corresponds to the upper level of inflation the Swiss National Bank is willing to tolerate according to their definition of price stability. With regards to the minimum coverage ratio CR_{min}, two different levels will be considered. The first setup sets CR_{min} to 90% which corresponds to a typical private pension fund setup and the second setup sets CR_{min} to 80% which reflects the setup option available to a government-related entity. The table below shows the simulation results for the base case . In each row the * and ° symbols indicate that the mean is significantly higher compared to the next lower mean at the 1% respectively 5% significance level[27]. Thereby, a two-sided t-test for the difference in means with unknown variances is conducted in every line except for the ERCPerc risk measures where a bootstrap approach is used to determine the significance[28]. The risk return metrics show the means of the annually compounded returns, the mean annual standard deviations as well as the average Sharpe ratios of all simulated paths for the six strategies introduced in section 3.3. The highest return and risk is related to the full equity investment (FEQ), followed by the put option replication strategy (PORS), and the stop-loss strategy (SLS)[29].

[27]For the employer risk measures the * and ° symbols indicate that the mean is significantly lower compared to the next higher mean at the 1% respectively 5% significance level.

[28]In the bootstrap approach for every strategy 1000 bootstrap samples are drawn from the 5000 ERC samples obtained from the Monte Carlo simulation. For every bootstrap sample the ERCPerc value is then calculated for $p = 5\%$, which gives a distribution of the ERCPerc5% metric for every strategy. These distributions are then used in a two-sided t-test for the difference with unknown variances in order to determine whether the mean ERCPerc5% level between two strategies are significantly different.

[29]In the 80% coverage ratio case SLS shows a higher mean return compared to PORS, but the difference is not significant at the 5% significance level.

Table 3.1: Base Case Scenario

The table shows the results of a Monte Carlo simulation with 5000 runs for the base case scenario. Two minimum coverage ratio settings are considered (90% and 80%). For each coverage ratio setting and investment strategy the risk return metrics, as well as the employee and employer preferences as specified in subsection 3.2.2 are obtained. In each row, except for the employer preferences, the * and ° symbols indicate that the mean is significantly higher compared to the next lower mean at the 1% respectively 5% significance level. For the employer risk measures the * and ° symbols indicate that the mean is significantly lower compared to the next higher mean at the 1% respectively 5% significance level. Thereby, a two-sided t-test for the difference in means with unknown variances is conducted in every line except for the ERCPerc5% risk measures where a bootstrap approach is used to determine the significance.

	CASH	**FEQ**	**50/50**	**CPPI**	**PORS**	**SLS**
Min Coverage Ratio 90%						
Risk Return Metrics						
Mean Annual Compounded Return	4.0%	7.5%*	6.1%*	5.9%*	6.9%*	6.6%*
Mean Annual Standard Deviation	2.2%	16.9%*	8.7%*	11.7%*	14.5%*	15.6%*
Mean Sharpe Ratio	0.00	0.21*	0.25*	0.14*	0.19*	0.16*
Employee						
AggCRRA	18.9	51.5*	40.2*	36.8*	45.7	45.6*
AggPV	22.2	137.7*	62.4*	71.4*	103.8*	109.6
Employer						
ERC	−15.2	−6128.1*	−460.1*	−333.4*	−3239.4*	−5881.8*
ERCPerc5%	0.0	−13631.7*	−2262.6*	−1306.9*	−6666.5*	−10446.4*
Min Coverage Ratio 80%						
Risk Return Metrics						
Mean Annual Compounded Return	4.0%	7.5%*	6.1%*	6.5%*	7.2%*	7.3%
Mean Annual Standard Deviation	2.2%	16.9%*	8.7%*	13.6%*	15.8%*	16.5%*
Mean Sharpe Ratio	0.00	0.21*	0.25*	0.16*	0.20	0.20*
Employee						
AggCRRA	18.9	50.7*	39.9*	41.4*	48.1*	49.5*
AggPV	22.3	132.9	61.8*	91.2*	117.4*	126.7*
Employer						
ERC	−4.1	−4571.5*	−220.6*	−916.8*	−3592.3*	−5466.2*
ERCPerc5%	0.0	−11732.6*	−1895.5*	−3348.0*	−8692.6*	−11885.3*

Then comes the 50/50 strategy closely followed by the CPPI strategy in the 90% minimum coverage ratio setting while it is the other way around in the 80% minimum coverage ratio setting. The CASH strategy shows the lowest returns. With respect to the Sharpe ratio, the dynamic strategies have lower values than the passive FEQ and 50/50 strategies for both coverage level settings. The CPPI strategy exhibits the lowest Sharpe ratio. With respect to the three dynamic strategies it is noteworthy, that they exhibit higher returns (and risk) when the coverage ratio is lower. This can be expected by the lower floor level which in general leads to a higher allocation in the risky asset over time, for all three dynamic strategies.

When moving to the perspective of the employees, the values of the different preference specifications are shown as defined in equations 3.14 to 3.16 in section 3.2.1. As it is seen in table 3.1, the order of the preferences for the different strategies depends on the choice of the utility specification. With regards to the CRRA specification, FEQ is the preferred choice, followed by PORS, SLS, 50/50 and CPPI when an 90% minimum coverage ratio is assumed[30]. The lowest utility is attributed to the full cash investment (CASH). For the prospect theory value (PV) specification the preferences for the 90% coverage ratio level are as follows: FEQ is the most preferred strategy followed by SLS, PORS, CPPI, 50/50 and CASH. Thus, when prospect theory preferences are assumed, the CPPI approach is preferred over the balanced 50/50 strategy which is preferred over the CPPI approach under expected utility theory (CRRA). Under the 80% minimum coverage ratio setting, the preference ordering is the same for the PV and CRRA specifications. It is: FEQ, SLS, PORS, CPPI, 50/50 and CASH.

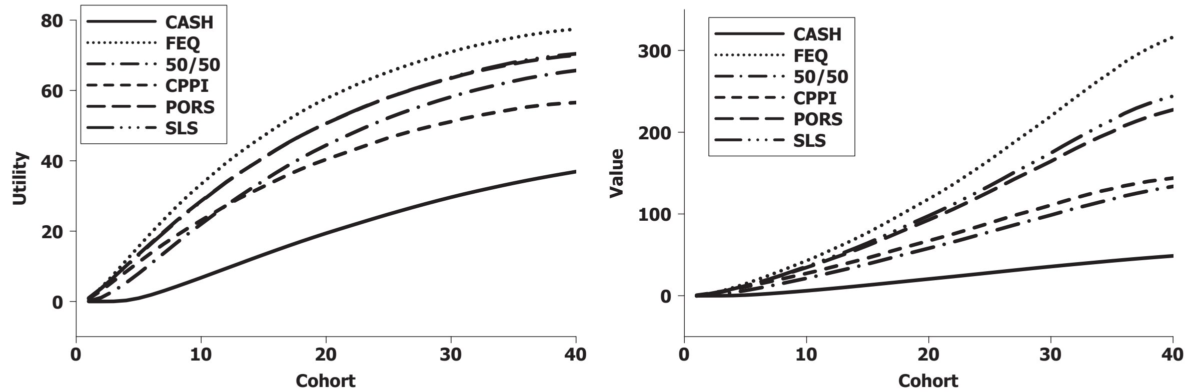
The result that the most risky strategy FEQ is the preferred choice under all preference specifications may surprise at first sight, as typical pension funds in Switzerland have equity quotas which are significantly lower. The expected utility respectively the expected prospect value for every cohort

[30] The difference between PORS and SLS is not significant at the 5% level.

under the six strategies is depicted in figure 3.4. As it is seen, the FEQ strategy is preferred by most cohorts except the oldest ones which are very close to retirement[31]. However, it needs to be recalled at this point, that the employees do not carry the full downside risk of the strategies as the employer is partially liable for any funding gaps. Thus, an analysis of the preferences of the employer is crucial.

Figure 3.4: Value and Utility per Cohort - Base Case Scenario

The graphs show the average CRRA utility $\frac{1}{M}\sum_{m=1}^{M} CRRA(\Delta x_{y,m})$ (left graph) respectively the average prospect value $\frac{1}{M}\sum_{m=1}^{M} v(\Delta x_{y,m})$ (right graph) of every cohort y for the six strategies under consideration (CASH, FEQ, 50/50, CPPI, PORS and SLS). For the analysis a 2% fixed rate is used and equal demographic weighting is applied. The 90% minimum coverage ratio setting is applied. The utility respectively the prospect value of the cohorts for every strategy is referenced on the vertical axis.



As it is explained in section 3.2.2, the preference of the employer are measured by the expected restructuring costs respectively the expected restructuring cost percentile as defined in equations 3.17 and 3.18. Since the employer has no upside participation when a risky strategy is chosen, the employer preferences are almost a mirror image of the employees preferences as shown in table 3.1. When the minimum coverage level is chosen at 90%, the preference ordering from the employer's view is: CASH, CPPI, 50/50, PORS, SLS and FEQ. When the minimum coverage level is set to 80% the ordering is: CASH, 50/50, CPPI, PORS, FEQ and SLS. Thus, the less risky strategies are naturally preferred by the employer and the ordering depends

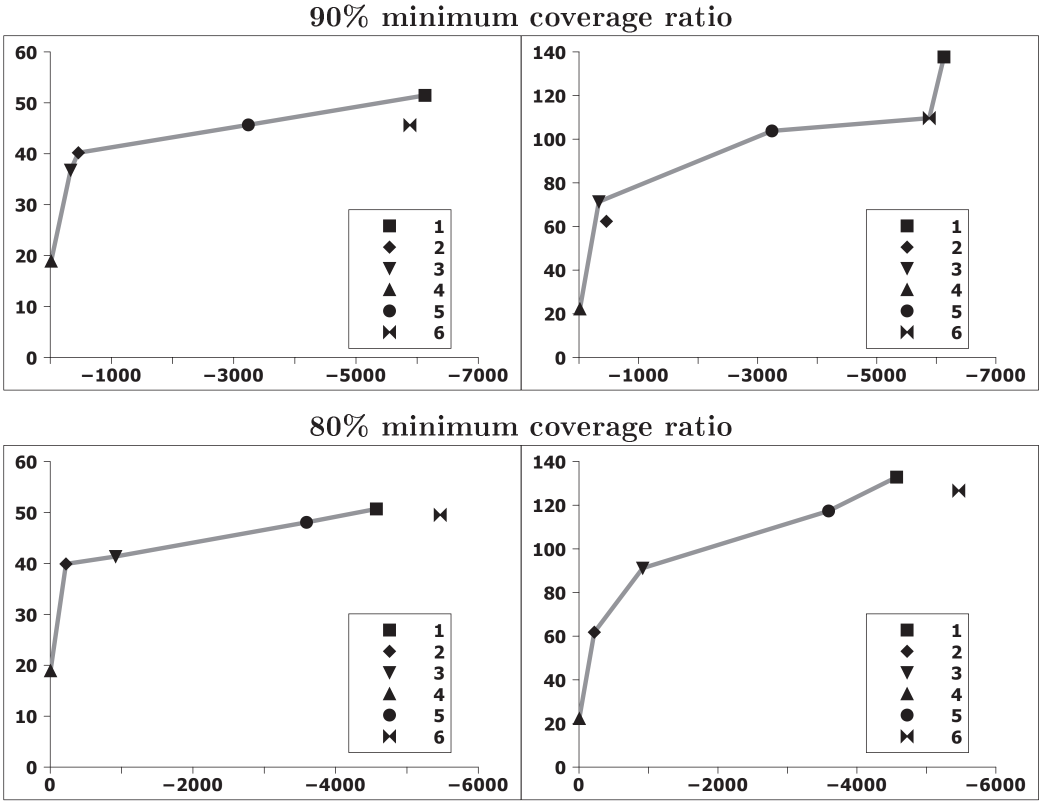
[31]For the CRRA specification only cohort 1 does not prefer FEQ over the other strategies. Under a PV specification it is only cohorts 1 and 2 which do not prefer FEQ over the other strategies.

on the choice of the coverage ratio.

Given the employer's as well as the aggregate employee's views, the question is which strategy should be chosen by a pension fund[32]. As the results indicate, the preferences of the employer and the aggregate employees are to a large extent contrary to each other. Figure 3.5 plots the employee preferences (vertical axis) versus the employer preferences (horizontal axis) for the two employee preference function specifications. As it is shown in figure 3.5, depending on the minimum coverage ratio setting as well as the preference specification of the aggregate employees, certain strategies may actually be dominated by others. Thereby, a certain strategy A is dominated by another strategy B when strategy B exhibits at least as much or more value to the employer as well as to the employee as does strategy A. Strategies which are not dominated by another strategy are connected by a line in the figure below, constituting an efficient strategy frontier.

[32]Within Swiss pension funds often a bipartisan committee with representatives from the employee's and the employer's side decides on the broad parameters of the investment strategy.

Figure 3.5: Efficient Strategy Frontier - Base Case Scenario
The following graphs show the simulation results of the base case scenario for a minimum coverage ratio of 90% (top row) respectively 80% (bottom row) for two different employee preference specifications. The vertical axis shows the aggregate employee preferences measured by the CRRA and PV specification respectively (from left to right). The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.



The results in figure 3.5 show, that in the 80% minimum coverage ratio setup, the stop loss strategy (SLS, number 6) is dominated by a full equity investment (FEQ, number 1). Put differently, the FEQ strategy yields higher value to the employees, while providing for lower expected restructuring contributions for the employer. Thus both, the employer and the employees, would prefer strategy FEQ over SLS. In the 90% coverage ratio setup the situation is marginally different. Under CRRA preferences, again the SLS strategy is dominated by FEQ. However, when analyzing the prospect theory specifications under the 90% minimum coverage ratio regime, the results indicate, that the 50/50 approach is dominated by the CPPI strategy.

This result is crucial as it suggests that it is possible that a dynamic approach is the preferred choice over a typical semi-passive 50/50 strategy which can be seen as a proxy of actual strategies deployed by pension funds. In order to understand how this result is obtained, the utility respectively the prospect value of each cohort for the two strategies need to be investigated. As it is seen in the left graph of figure 3.6, under CRRA preferences only the older cohorts prefer the CPPI strategy over the 50/50 strategy. From cohort 12 onwards, the 50/50 strategy is preferred over the CPPI strategy, which is indicated by the crossing of the two solid lines. This is interesting, as the mean excess return of the CPPI strategy for the cohorts is higher for every cohort under consideration (dotted lines). However, the median excess return is lower for the CPPI strategy (dashed lines). Under PV preferences, shown in the right graph of figure 3.6, every cohort prefers the CPPI strategy over the 50/50 strategy. Thus, the distributional properties of the CPPI strategy are disadvantageous relative to the distribution of the 50/50 strategy when the concavity of the preference function is high as is is the case under the CRRA specification[33]. To get an idea of the distributional properties of the two strategies, figure 3.7 shows the excess return distribution (Δx) of cohort 10 for the 50/50 and CPPI strategy. As it can be seen, the CPPI strategy has much more mass around 0 and a thinner left tail compared to the 50/50 strategy. Further, also extreme positive returns are more likely under a CPPI strategy compared to the 50/50 approach. The thin left tail of the distribution obtained by using a CPPI strategy also explains the preference for the CPPI strategy relative to the 50/50 strategy from an employers point of view in the base case scenario. Looking at the relative contribution of each cohort in the two lower graphs in figure 3.6, it can be seen that the younger cohorts, have a slightly disproportionate effect on the aggregate preference results for the CRRA as well as for the PV specification. As it is shown in the sensitivity analysis in the next subsection, the results concerning the dominance of certain strategies are sensitive to variations in the parameter

[33]See figure 3.2 for a comparison of the CRRA and PV function shapes.

specifications, in particular with respect to the fixed rate r_{fix}.

Figure 3.6: Per Cohort Contributions (CPPI versus 50/50) - Base Case Scenario

The graphs in the top row show the average CRRA utility $\frac{1}{M}\sum_{m=1}^{M} CRRA(\Delta x_{y,m})$ (left graph) respectively the average prospect value $\frac{1}{M}\sum_{m=1}^{M} v(\Delta x_{y,m})$ (right graph) of every cohort y for the 50/50 as well as the CPPI strategy. For the analysis a 2% fixed rate is used and equal demographic weighting is applied in a 90% minimum coverage ratio setup. The cohorts are ordered along the horizontal axis from the oldest to the youngest cohort. The utility respectively the prospect value of the cohorts for the two strategies is referenced on the vertical axis on the left hand side. In addition to the utility respectively value lines, also the mean and median excess return of the two strategies for every cohort are shown. The excess returns are referenced on the vertical axis on the right hand side. The two lower graphs visualize the preference contribution of each cohort with respect to the two strategies. Thereby, the bars represent the difference between the utility/value of the CPPI strategy minus the utility/value of the 50/50 strategy for a specific cohort.

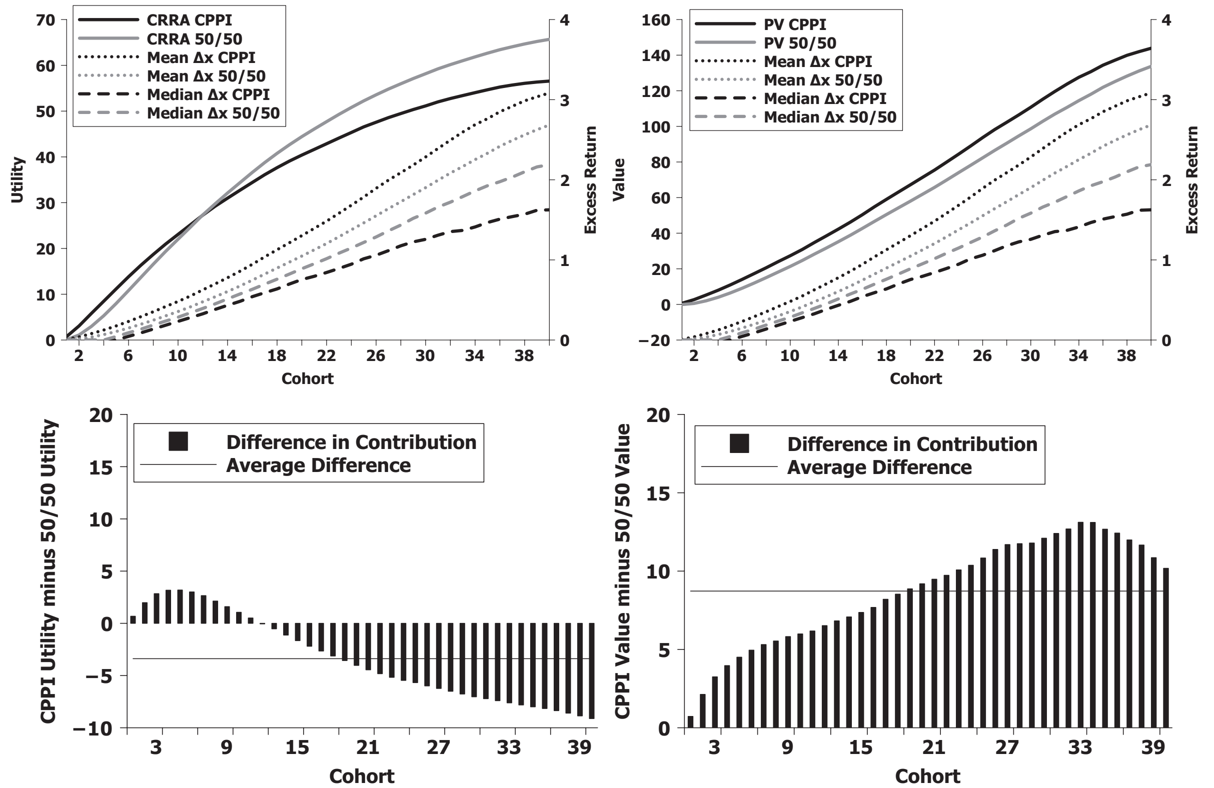
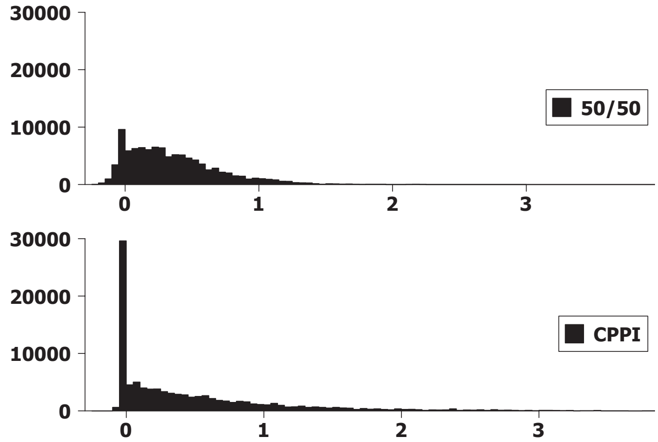


Figure 3.7: Excess Return Distribution (CPPI versus 50/50) - Base Case Scenario

Distribution of Δx for cohort 10 given the 50/50 strategy (upper graph) and the CPPI strategy (lower graph) in the base case scenario.

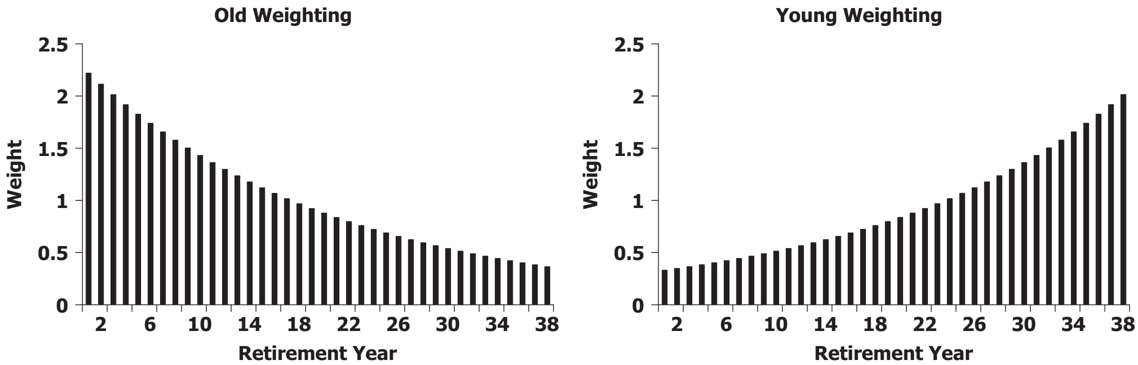


3.4.2 Sensitivity Analysis

This section shows the main sensitivity results with respect to changes in the assumptions made in the parameter specifications. The focus will be on two critical parameters. The first one is the fixed rate r_{fix} which, as will be shown, has a crucial impact on the relative attractiveness of the strategies in the given Pension fund framework. In the following sensitivity analysis the fixed rate is varied from 0% to 5% in 1% increments. The second important parameter is the size of the cohorts respectively the relative weighting of the different cohorts. The weighting factor d_y, introduced in section 3.2.2 allows to model different assumptions about the demographic distribution within a pension fund. By applying different weightings to the different cohorts it is for example possible to model a pension fund setting where there are significantly more older employees which will retire earlier than there are younger employees which will retire later. According to the results of Dierkes, Erner and Zeisberger (2010), the investment horizon is crucial for the preferences of prospect theory investors, therefore the influence of different demographic settings may be of importance. In order to show the influence of the demographic assumptions, two extreme cases are compared to an equally weighed setting as it was used in the base case scenario. The first extreme case assumes a strong overweight of older cohorts. In particular, it is assumed that cohorts shrink by 5% when compared to the previous cohort which is one year closer to retirement. The second extreme case assumes the contrary, namely that each cohort is 5% larger compared to the cohort which retires one year earlier. A plot of the two weighting functions is given in figure 3.8:

Figure 3.8: Demographic Weighting Functions

The graphs show the demographic weighting functions applied in the sensitivity analysis. Thereby, the old demographic weighting scheme assumed that cohorts shrink by 5% when compared to the previous cohort which is one year closer to retirement (left graph). On the other side, the young demographic weighting scheme assumes that each cohort is 5% larger compared to the cohort which retires one year earlier (right graph).



The following figures show the fixed rate sensitivity for the two different employee preferences specifications as well as the two minimum coverage ratio specifications for the three demographic weighing schemes. Figures 3.9 and 3.10, display the results when equal weighted cohorts are assumed for the CRRA and PV employee preference specification respectively.

Figure 3.9: Fixed Rate Sensitivity Analysis - Equal Demographic Weighting Base Case Scenario CRRA Preferences

The following graphs show the simulation results of the sensitivity analysis assuming equal demographic weighting for a minimum coverage ratio of 90% respectively 80% for the CRRA employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee utility measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)

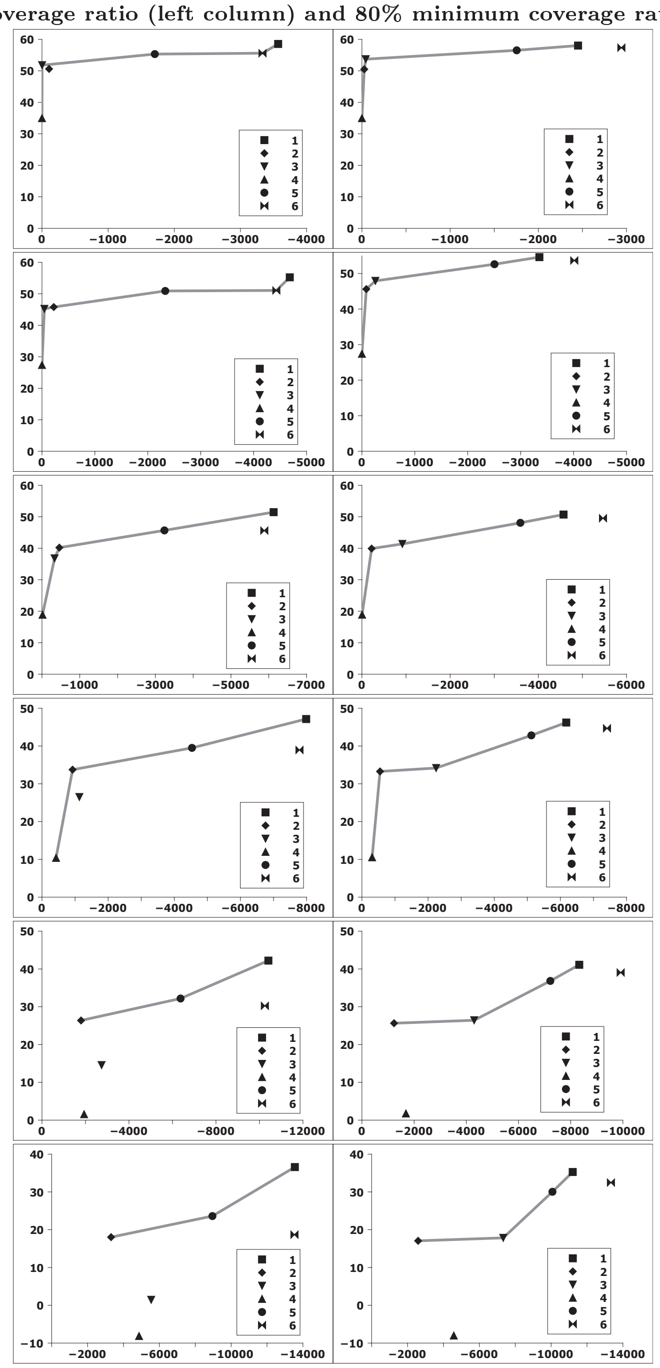
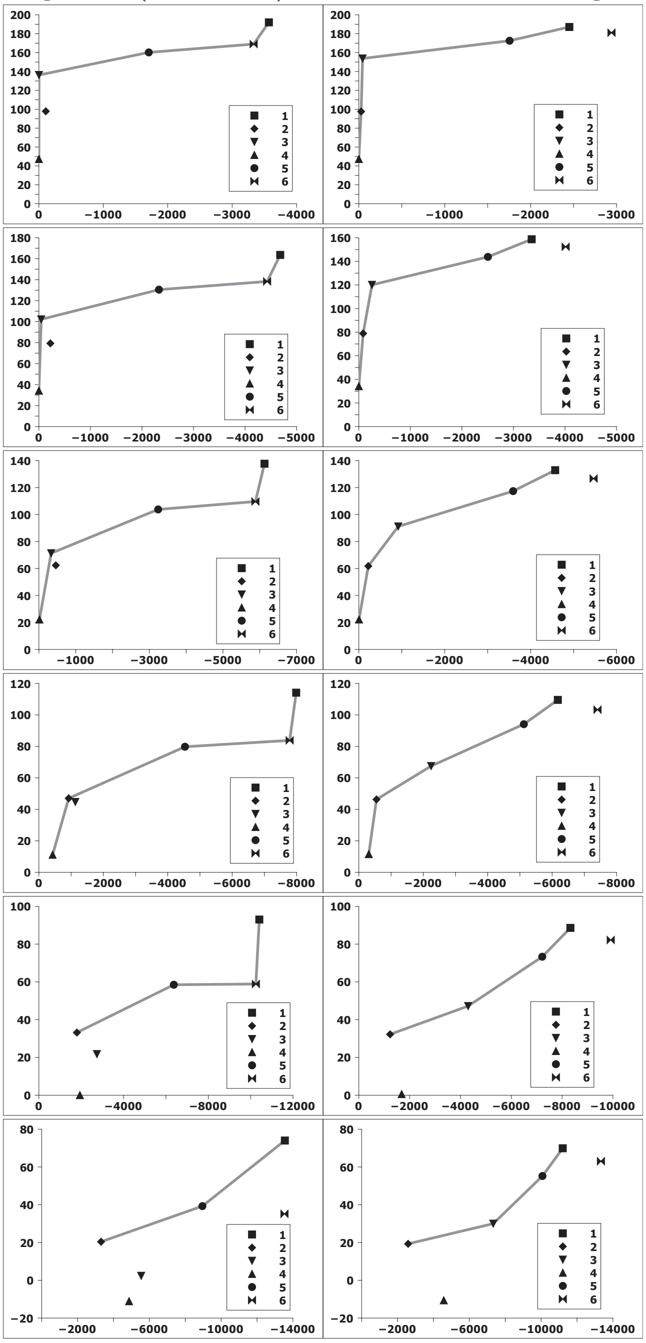


Figure 3.10: Fixed Rate Sensitivity Analysis - Equal Demographic Weighting Base Case Scenario PV Preferences

The following graphs show the simulation results of the sensitivity analysis assuming equal demographic weighting for a minimum coverage ratio of 90% respectively 80% for the PV employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee value measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)



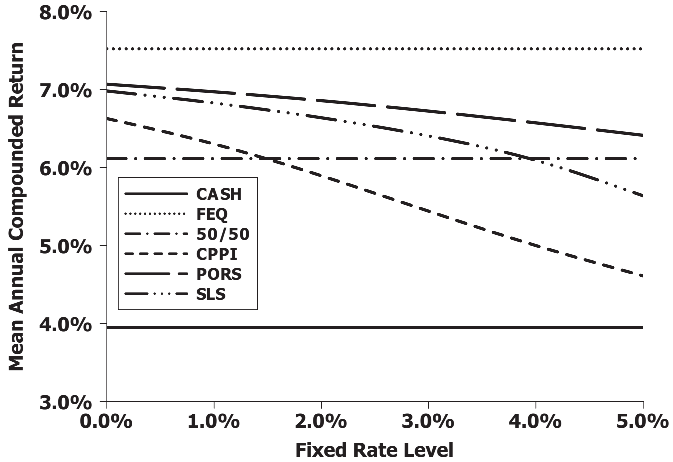
The sensitivity analysis with respect to the choice of the fixed rate as shown in figures 3.9 and 3.10, reveals that independent of the employee preference specification (CRRA or PV), a higher fixed rate makes the more aggressive strategies (FEQ=1, PORS=5) more attractive relative to the other strategies from an aggregated employee's point of view. This is reflected by the steepening of the strategy efficient frontier when moving from a 0% fixed rate in the top plots to the 5% fixed rate in the bottom plots. This observation is to be expected, as the likelihood of achieving the high fixed rate returns with defensive strategies is smaller, compared to the more aggressive strategies. When analyzing the different strategies in detail, it is observed that the full equity allocation FEQ has the highest aggregated employee value independent of the utility specification as well as the choice of the fixed rate. However, if the fixed rate is low, the additional value to employees when choosing the most aggressive strategy is only modest compared to the other strategies. The stop-loss strategy (SLS) in general performs relatively poorly. In the 80% minimum coverage ratio setting, SLS is always dominated by FEQ. In the 90% minimum coverage ratio setup SLS is only dominated by FEQ if a high fixed rate is chosen, however at lower fixed rates FEQ yields considerably more value to the aggregated employees value function compared to SLS while only being slightly more risky for the employer.

As is seen in figures 3.9 and 3.10, the PORS strategy is never dominated by another strategy and basically represents a less risky alternative to the FEQ approach from an employer's point of view. A full cash investment (CASH) always yields the lowest value to the employees. It is dominated by CPPI under different settings, when the fixed rate is chosen at 0%. CASH is also dominated whenever the fixed rate is equal to or higher than the assumed long term risk free rate of 4%.

Further, the sensitivity analysis shows an interesting interaction between the CPPI and 50/50 strategies. In the 80% minimum coverage ratio setting the CPPI strategy exhibits a higher aggregated employee value compared to the 50/50 strategy in all considered cases, independent of the preference specification. At the same time in this setting the expected employer contribution are always higher for the CPPI approach. However, the ratio between the employees preferences and the employer contributions changes considerably as the fixed rate is increased (moving from the top to the bottom in figures 3.9 and 3.10) in favor of the 50/50 strategy. Also, in general this ratio is more favorable with respect to the CPPI strategy under the PV specifications. In the 90% minimum coverage ratio setting, the CPPI strategy dominates the 50/50 strategy when the fixed rate lies between 0% and 2% and when PV preferences are assumed. For the CRRA preference specification, CPPI dominates 50/50 only when the fixed rate is set at 0%. Once the fixed rate is chosen at 3-5%, the 50/50 strategy dominates CPPI, under both preference specifications. Thus, in the 90% minimum coverage ratio setting there is a sliding schedule of the preferences and dominance between the CPPI and 50/50 strategies, moving in favor of 50/50 as the fix rate is increased. These results reveal that the aggregate employee specification and the choice of the fixed rate are crucial elements in determining which strategy is optimal in a given pension fund setting. In particular the attractiveness of the CPPI strategy depends on the difference between the fixed rate and the assumed long term risk free rate. As the fixed rate approaches the long term risk free rate, which is assumed to be 4% in the base case, CPPI becomes less attractive versus the 50/50 strategy and vice versa[34]. One reason for this is the fact, that the expected return of the CPPI strategy (or any dynamic strategy) decreases as the average strike level is implicitly increased by choosing a higher fixed rate. At the same time the expected annual return for the 50/50 strategy is independent of the fixed rate. The mean compounded annual returns of all strategies as a function of the fixed minimum rate are shown in figure 3.11.

[34]The relationship between the fixed rate and the average interest rate will be further investigated in the current environment case study in subsection 3.4.3.

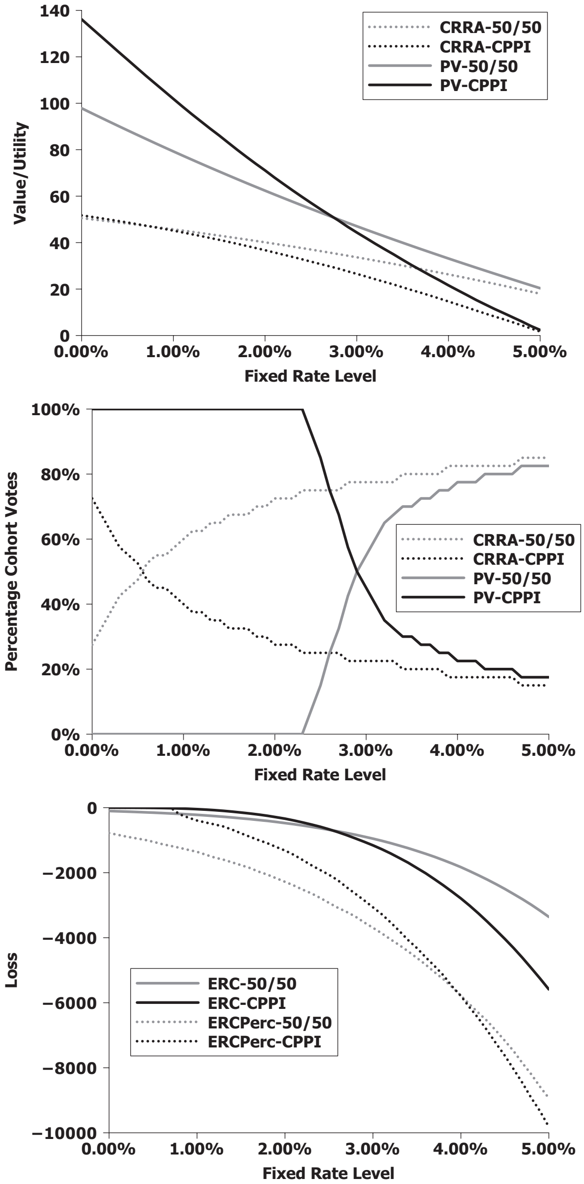
Figure 3.11: Mean Annual Returns as a Function of the Fixed Rate - Base Case Scenario
This graph depicts the mean annual compounded returns of the CASH, FEQ, 50/50, CPPI, PORS and SLS strategies on the vertical axis as a function of the fixed rate on the horizontal axis. Equal demographic weighting and a 90% minimum coverage ratio setting are applied.



In order to get a better understanding of the relative attractiveness of the CPPI and 50/50 strategies, figure 3.12 shows how the preferences for the two strategies changes as a function of the fixed rate level. Thereby, the top graph shows the aggregate utility respectively value of the two strategies as a function of the fixed rate level. The point where the two lines cross can be interpreted as a break-even rate, where the employee is indifferent between the two strategies. Under CRRA preferences the break-even rate is found at a fixed rate of 0.7% while under PV preferences the break-even rate is at 2.7%. As it was shown in the bottom graphs of figure 3.6, the younger cohorts contribute disproportionately to the aggregate preference measure. Thus, it may be the case that one strategy is preferred based on the aggregate preference measure, while in fact the majority of the cohorts would prefer the other strategy. In order to analyze this effect, the middle graph shows the preferences for the two strategies based on percentage of cohorts which would vote in favor of a certain strategy at a given fixed rate (voting preferences). The corresponding break even rates for the CRRA and PV preferences are 0.6% respectively 2.9% respectively. Thus, the contortion of aggregating the preferences relative to the view from a voting perspective is small. The bottom graph in figure 3.12 depicts the preferences from the employer's point of view, plotting the ERC and ERCPerc preferences for the two strategies as a function of the fixed rate. The corresponding break-even rates from an employers perspective are 2.6% under ERC preferences and 4% under ERCPerc preferences. As the analysis of the dominance limits has revealed, the dominance of CPPI over 50/50 depends strongly on the preference measures used for the employer as well as the employee. The minimum break-even rate is 0.6% when CRRA voting preferences are used and the maximum break-even rate is 2.9% obtained by using PV voting preferences for the employees and the ERCPerc preference measure for the employer.

Figure 3.12: Dominance Limits - Equal Demographic Weighting Base Case Scenario

The following graphs show the aggregate employee preferences (top), the employee voting preferences (middle) as well as the employer preferences (bottom) as a function of the fixed rate for the 50/50 as well as the CPPI strategy. The aggregate employee preferences correspond to the CRRA and PV specifications as defined in equations 3.14 to 3.16. The employee voting preferences depict the percentage of cohorts for which the specified strategy is preferred given the fixed rate. The employer preferences correspond to the ERC and ERCPerc risk measures as defined in equations 3.17 to 3.18.



With respect to the demographic weighing schemes analyzed in figures 3.13, 3.14, 3.15 and 3.16, it can be seen that in general, the demographic distribution does not have a large impact on the overall results.

Figure 3.13: Fixed Rate Sensitivity Analysis - Old Demographic Weighting Base Case Scenario CRRA Preferences

The following graphs show the simulation results of the sensitivity analysis assuming an overweighting of old cohorts for a minimum coverage ratio of 90% respectively 80% for the CRRA employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee value measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)

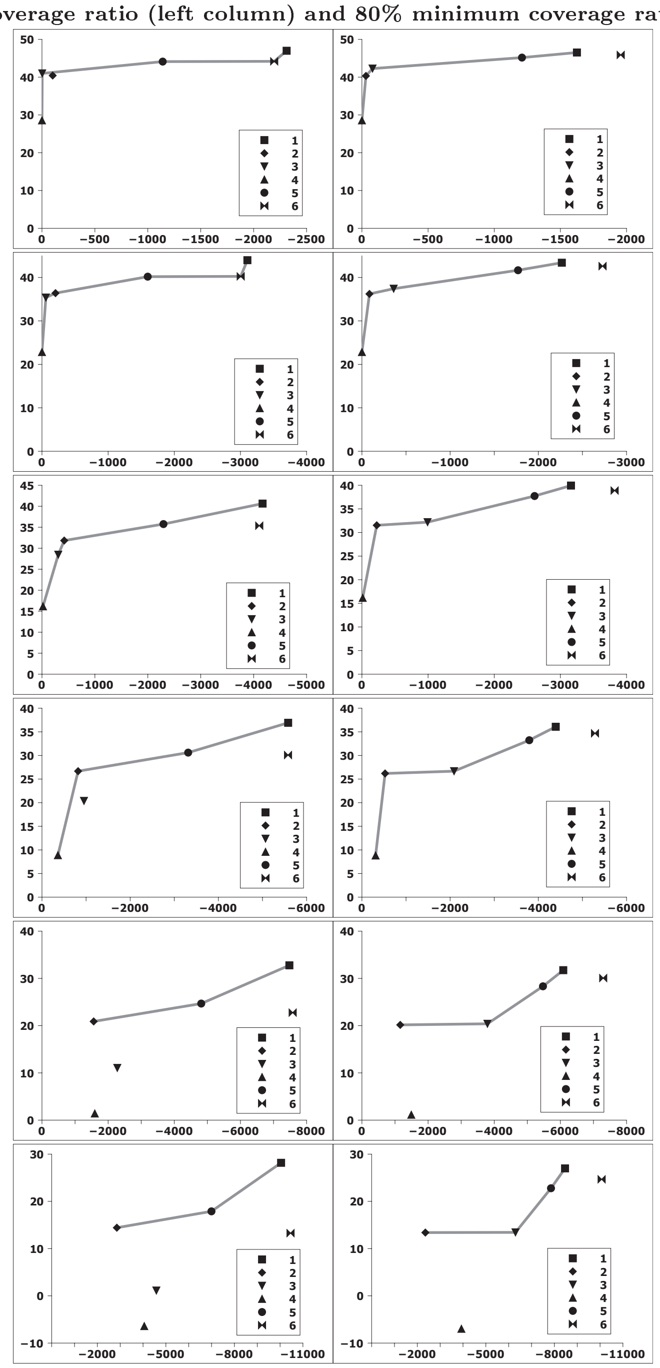


Figure 3.14: Fixed Rate Sensitivity Analysis - Old Demographic Weighting Base Case Scenario PV Preferences

The following graphs show the simulation results of the sensitivity analysis assuming an over-weighting of old cohorts for a minimum coverage ratio of 90% respectively 80% for the PV employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee value measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)

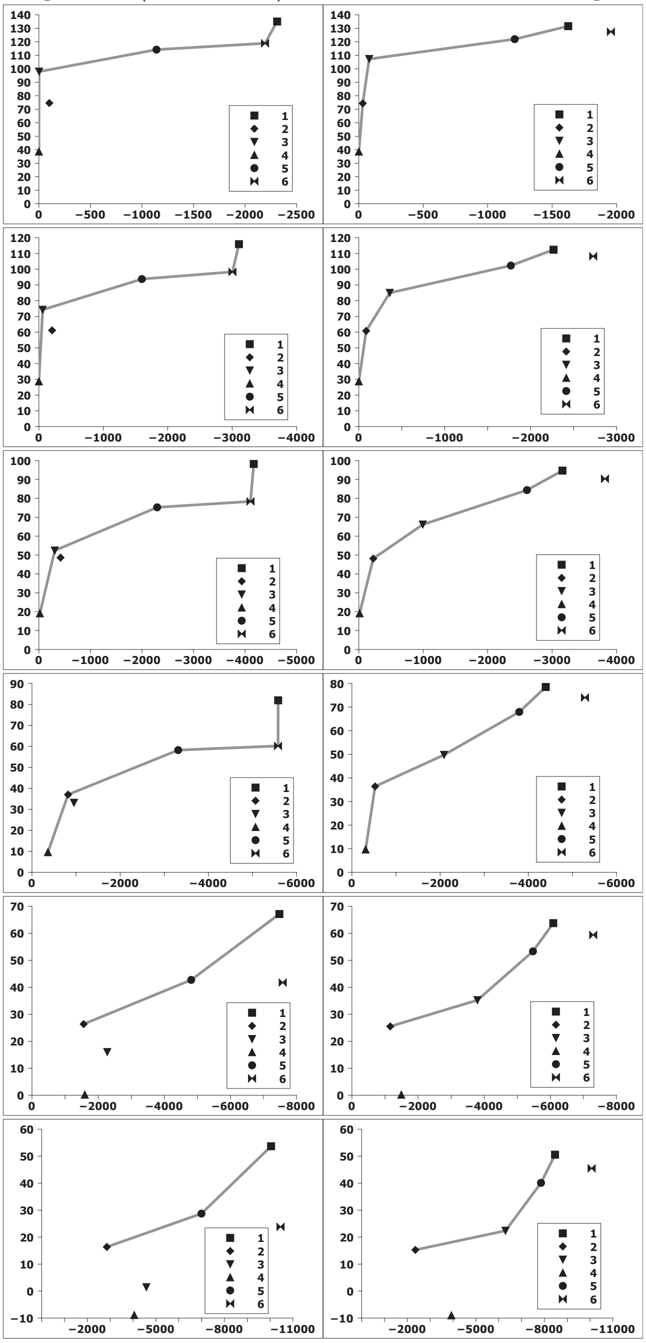


Figure 3.15: Fixed Rate Sensitivity Analysis - Young Demographic Weighting Base Case Scenario CRRA Preferences

The following graphs show the simulation results of the sensitivity analysis assuming an over-weighting of young cohorts for a minimum coverage ratio of 90% respectively 80% for the CRRA employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee value measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)

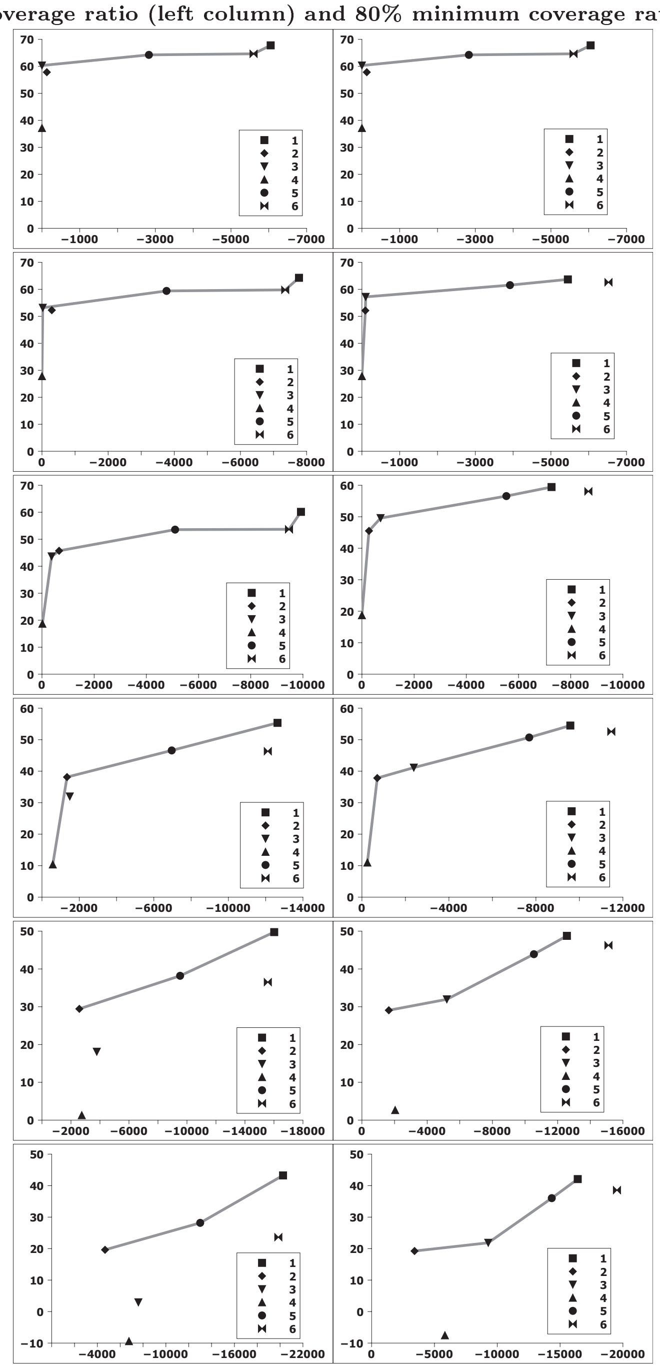
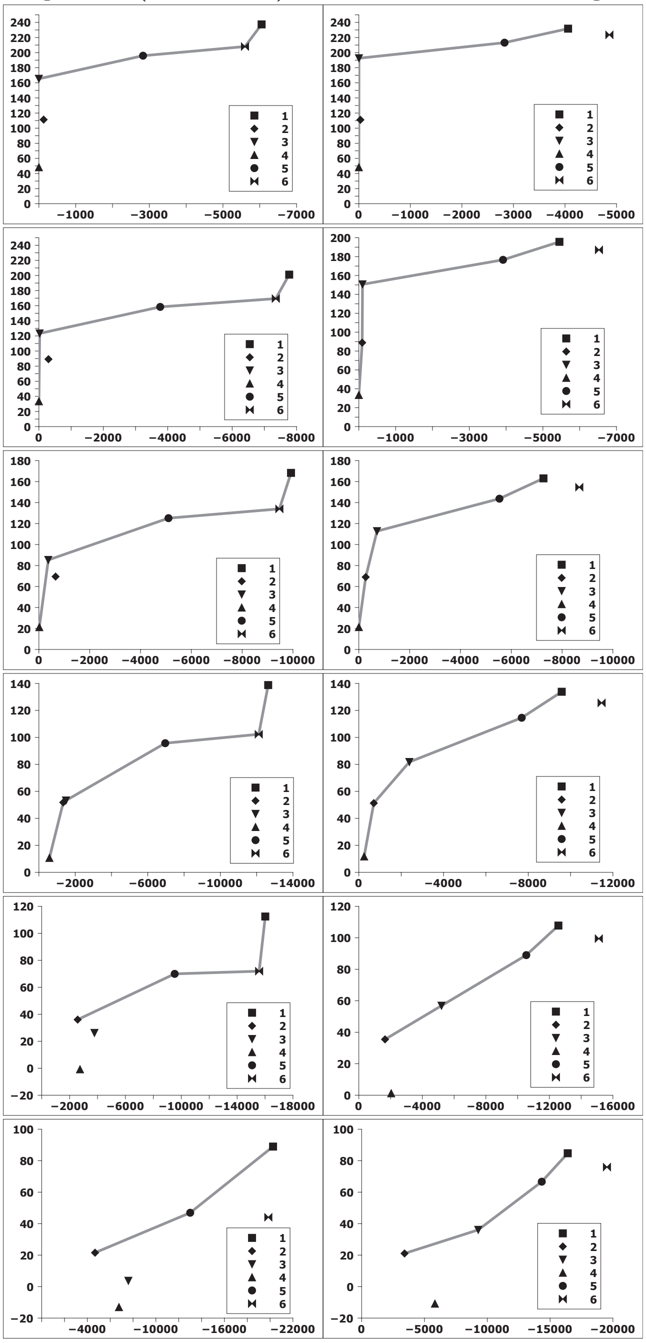


Figure 3.16: Fixed Rate Sensitivity Analysis - Young Demographic Weighting Base Case Scenario PV Preferences

The following graphs show the simulation results of the sensitivity analysis assuming an over-weighting of young cohorts for a minimum coverage ratio of 90% respectively 80% for the PV employee preference specification under different choices for the fixed rate. In each column, the results are shown for the fixed rate varying from 0% to 5% in 1% increments from the top to the bottom. The vertical axis of each plot shows the aggregate employee value measured by the relevant preference specification. The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio (left column) and 80% minimum coverage ratio (right column)



A noteworthy small effect is however observed with respect to the relationship between the CPPI and 50/50 strategy. As the results indicate, CPPI is more attractive from an aggregate employees point of view relative to the 50/50 strategy when the young cohorts are over-weighted relative to the older cohorts. This is seen for example in the 90% minimum coverage ratio 1% CRRA specification in tables 3.13 and 3.15 or the 90% minimum coverage ratio 3% PV specification in tables 3.14 and 3.16. In the first case, CPPI dominates the 50/50 strategy under the young demographic weighting scheme, while it does not under the old demographic weighting scheme. In the second case, CPPI is dominated by the 50/50 strategy under the old scheme while this dominance disappears in the young scheme.

In order to see the demographic weighting effects with respect to the preferences for the CPPI and 50/50 strategy, again the dominance limits are analyzed as shown in figure 3.17. In the left row the break-even rates for the old-demographic weighting scheme are shown. As can be seen with respect to the employee aggregate preferences, the break-even rates are lower relative to the equal weight case shown in figure 3.12. The break-even rates for the aggregate preferences in the top graph are 0.4% and 2.4% for the CRRA and PV specification respectively. The voting preferences however are actually slightly higher at 0.8% and 3% respectively. This result is attributable to the circumstance that the voting weights of the older cohorts, which prefer the CPPI strategy over the 50/50 strategy, are over-weighted[35]. Moving to the break-even rates for the case where the young cohorts are over-weighted reveals, that the break-even rates for the aggregate preferences are higher relative to the equal weighted case. The break-even rates are at 1.3% and 3.1% for the CRRA and PV specification respectively. The voting break-even rates for the young weighting scheme are 1.2% and 3.2% respectively. With respect to the employer preferences, the influence of the weighting scheme

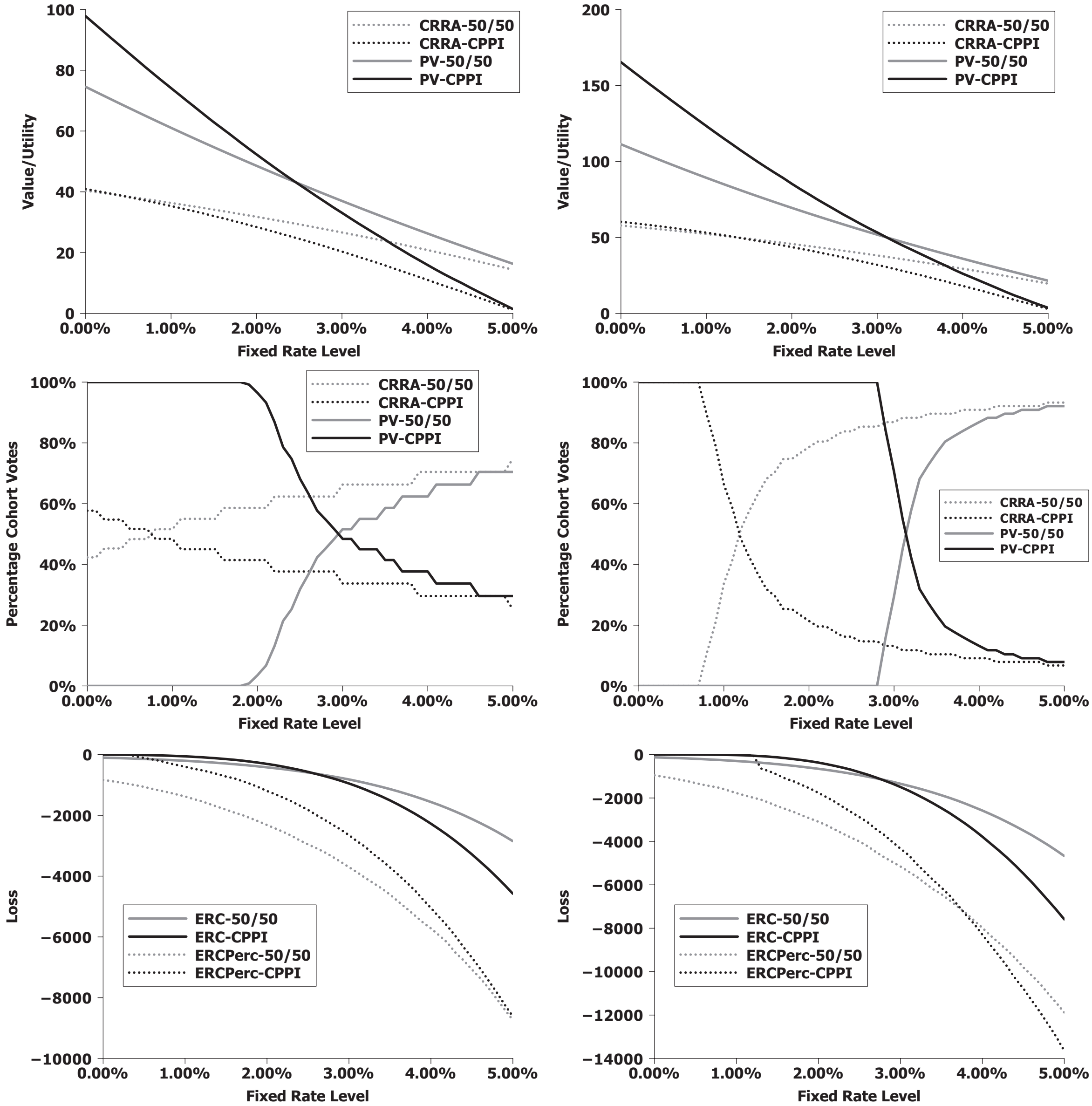
[35] Figure 3.25 in the appendix shows the per cohort contribution for the break-even rates under the equal weighting scheme. As it can be seen, the older cohorts prefer the CPPI strategy, while the younger cohorts are in favor of the 50/50 strategy.

on the preferences is less pronounced for the ERC preference measure than the ERCPerc measure. The ERC break-even rates are at 2.6% and 2.8% for the old and young weighting scheme respectively. The corresponding ERCPerc break-even rates however stand at 5% and 3.8% for the old and young weighting scheme respectively.

In brief, the sensitivity analysis revealed that changes in the fixed rate may have a great impact on the relative attractiveness of the different strategies from an employee's as well as an employer's point of view. On the other side, changing the demographic weighting schemes only has a limited effect on the relative attractiveness of most strategies under consideration. However, with respect to the dominance limits of the 50/50 and CPPI strategies, there may be a significant impact of the demographic weighting scheme on the relative attractiveness which may be of interest from a practical perspective. The results indicate that from an aggregate employee preference perspective the preference is shifted towards the 50/50 strategy under an old-demographic weighting scheme and vice versa under a young-demographic weighting scheme. From an employer's point of view, it is the other way around when the ERCPerc risk measure is applied while the weighting scheme has only a limited impact when the ERC measure is used. In the appendix, a detailed explanation is given about the origin of the preference shifts which occur when the demographic weighting is changed. In the next subsection two case studies as well as a simulation setup using a block-bootstrap approach are discussed in order to gain more insights from a practical perspective.

Figure 3.17: Dominance Limits - Old and Young Demographic Weighting Base Case Scenario

The following graphs show the aggregate employee preferences (top), the employee voting preferences (middle) as well as the employer preferences (bottom) as a function of the fixed rate for the 50/50 as well as the CPPI strategy. Thereby, the left column depicts the old-demographic weighting scheme and the right column depicts the young-demographic weighting scheme. The aggregate employee preferences correspond to the CRRA and PV specifications as defined in equations 3.14 to 3.16. The employee voting preferences depict the percentage of cohorts for which the specified strategy is preferred given the fixed rate. The employer preferences correspond to the ERC and ERCPerc risk measures as defined in equations 3.17 to 3.18.



3.4.3 Current Environment Scenarios

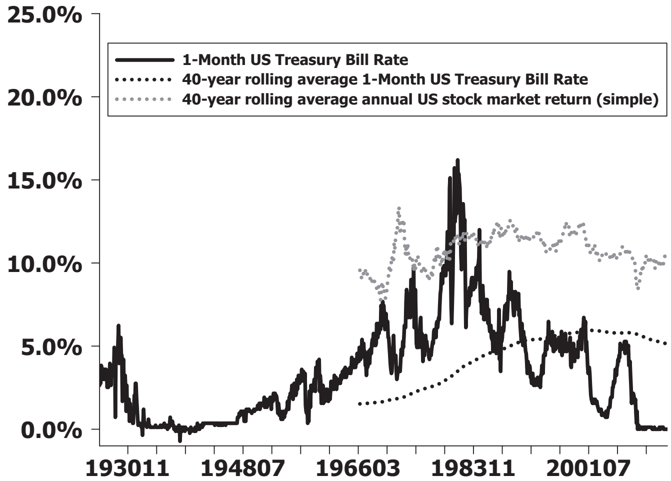
The base case scenario as it was described at the beginning of this section is based on the assumption that during the next 40 years, the risk free rate as well as the equity premium behave according to the long term observed average. As it was shown in the sensitivity analysis, the difference between the average risk free rate and the fixed rate is crucial for determining the preferences for each strategy. As the graph in figure 3.18 illustrates, the effective realized risk free rate over 40 years has actually varied significantly over time. Also, in the base case scenario it was assumed that the risk free rate at the beginning of the period is equal to the long term average of 4%. Given the current record low interest rates near 0% it is doubtful that any conclusions can be drawn for the current market environment. Therefore, in this subsection the base case scenario is adjusted in order to reflect the current market market environment and in order to gain insights into the role of long term interest rate expectations on the results. In particular, two extreme scenarios are analyzed. In the first scenario it is assumed that the risk free rate will stay very low on average for the next 40 years. This scenario is based on the period from 1930 until 1969 when the average monthly treasury bill rate in the US stood at 1.65% per annum[36]. The average annual return on the equity index was 11.18% (8.75% compounded) and the annual volatility was 22.45% for the same period[37]. The starting value for the risk free rate is set at 0.1%. This low rate scenario would reflect an environment where the loose monetary policy currently observed in many developed countries continues over a long period of time. The second scenario under consideration reflects a high interest rate period, where the chosen reference period spans the time from 1970 to 2009. The average monthly treasury bill rate stood at an annualized 5.6% for this period and the average expected annual return

[36] I choose the US market as the reference market since the longest consistent time series are available for this market.

[37] These numbers refer to the annual data series available on Kenneth French homepage: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

on equities was 11.52% (9.8% compounded). The annual volatility for this period stands at 18.65%. The starting value for the risk free rate is also set at 0.1%. The high rate scenario reflects an environment where interest rates are significantly above the long term average, possibly due to inflationary shocks which lead central banks to increase rates in order to stabilize prices. The pension funds under consideration are again one with a minimum coverage ratio of 90% and one with a minimum coverage ratio constraint of 80%. The fixed rate is again set at 2% as it was done in the base case, representing the long term expected inflation ceiling. Further, it is assumed that older cohorts are slightly over-weighted, whereby the population from one cohort to the next younger cohort decreases by 2%.

Figure 3.18: Historical Risk Free Interest Rate Levels and Equity Returns
The graph depicts the 1-month US Treasury bill rate for the period of July 1926 until August 2013. Further, it shows the 40-year rolling average 1-Month US Treasury Bill Rate as well as the 40-year rolling average compounded annual US stock market return. The data corresponds to the data used in Fama and French (1993).



The results for the low interest rate scenario are shown in table 3.2 as well as figure 3.19. Those for the high interest rate scenario are shown in table 3.3 respectively figure 3.20.

Table 3.2: Low Rate Scenario

The table shows the results of a Monte Carlo simulation with 5000 runs for the low rate case scenario. Two minimum coverage ratio settings are considered (90% and 80%). For each coverage ratio setting and investment strategy the risk return metrics, as well as the employee and employer preferences as specified in subsection 3.2.2 are obtained. In each row, except for the employer preferences, the * and ° symbols indicate that the mean is significantly higher compared to the next lower mean at the 1% respectively 5% significance level. For the employer risk measures the * and ° symbols indicate that the mean is significantly lower compared to the next higher mean at the 1% respectively 5% significance level. Thereby, a two-sided t-test for the difference in means with unknown variances is conducted in every line except for the ERCPerc5% risk measures where a bootstrap approach is used to determine the significance.

	CASH	**FEQ**	**50/50**	**CPPI**	**PORS**	**SLS**
Min Coverage Ratio 90%						
Risk Return Metrics						
Mean Annual Compounded Return	1.4%	8.5%*	5.6%*	4.0%*	6.6%*	6.0%*
Mean Annual Standard Deviation	1.3%	22.3%*	11.6%*	11.1%*	17.0%*	19.1%*
Mean Sharpe Ratio	0.00	0.32*	0.36*	0.17*	0.29*	0.22*
Employee						
AggCRRA	−4.5	52.9*	34.0*	19.5*	39.6	38.7*
AggPV	−7.6	218.2*	53.7*	39.1*	107.6*	119.6*
Employer						
ERC	2359.7	21726.7*	2355.3	3592.3*	8775.8*	17930.0*
ERCPerc5%	−4083.6	−51044.9*	−6669.5*	−6333.1*	−17594.1*	−34456.4*
Min Coverage Ratio 80%						
Risk Return Metrics						
Mean Annual Compounded Return	1.4%	8.5%*	5.6%*	5.9%*	7.5%*	7.8%*
Mean Annual Standard Deviation	1.3%	22.3%*	11.6%*	16.7%*	19.5%*	21.6%*
Mean Sharpe Ratio	0.00	0.32*	0.36*	0.24*	0.31*	0.29*
Employee						
CRRA	−4.1	52.2*	33.4*	35.1*	45.8*	49.5*
AggPV	−6.7	208.6*	52.3*	98.8*	149.0*	188.0*
Employer						
ERC	2128.2*	17541.1*	1598.4	8395.6*	12050.3*	21528.5*
ERCPerc5%	−3731.8	−43360.3*	−5728.4*	−18469.1*	−26930.9*	−46746.0*

Table 3.3: High Rate Scenario

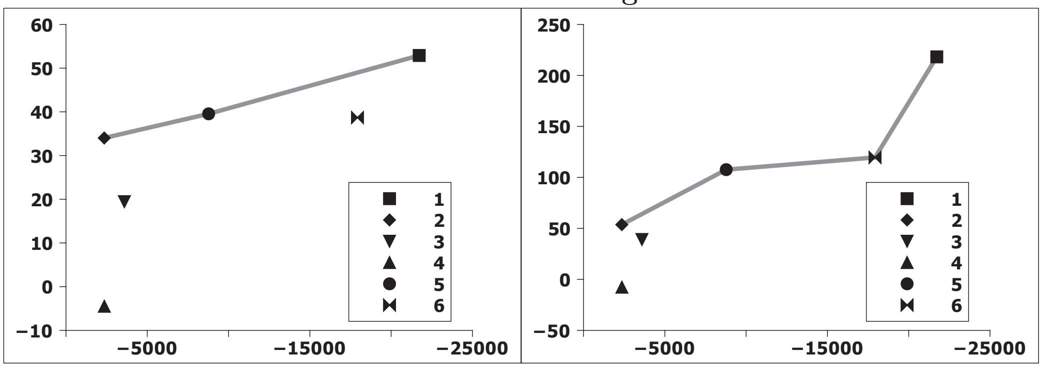
The table shows the results of a Monte Carlo simulation with 5000 runs for the high rate scenario. Two minimum coverage ratio settings are considered (90% and 80%). For each coverage ratio setting and investment strategy the risk return metrics, as well as the employee and employer preferences as specified in subsection 3.2.2 are obtained. In each row, except for the employer preferences, the * and ° symbols indicate that the mean is significantly higher compared to the next lower mean at the 1% respectively 5% significance level. For the employer risk measures the * and ° symbols indicate that the mean is significantly lower compared to the next higher mean at the 1% respectively 5% significance level. Thereby, a two-sided t-test for the difference in means with unknown variances is conducted in every line except for the ERCPerc5% risk measures where a bootstrap approach is used to determine the significance.

	CASH	**FEQ**	**50/50**	**CPPI**	**PORS**	**SLS**
Min Coverage Ratio 90%						
Risk Return Metrics						
Mean Annual Compounded Return	4.8%	9.7%*	7.7%*	7.7%	8.8%*	8.6%*
Mean Annual Standard Deviation	2.7%	18.5%*	9.6%*	13.8%*	16.1%*	17.3%*
Mean Sharpe Ratio	0.00	0.27*	0.30*	0.19*	0.24*	0.21*
Employee						
AggCRRA	20.2	55.9*	44.1*	40.5*	49.7	49.4*
AggPV	27.0	219.7*	91.0*	109.2*	159.4*	170.3*
Employer						
ERC	18.6	4940.4*	429.4*	169.8*	2371.3*	4751.0*
ERCPerc5%	0.0	−11530.3*	−2246.3*	−807.2*	−5102.2*	−8410.9*
Min Coverage Ratio 80%						
Risk Return Metrics						
Mean Annual Compounded Return	4.8%	9.7%*	7.7%*	8.4%*	9.3%*	9.3%
Mean Annual Standard Deviation	2.7%	18.5%*	9.6%*	15.5%*	17.4%*	18.2%*
Mean Sharpe Ratio	0.00	0.27*	0.30*	0.22*	0.25	0.25*
Employee						
AggCRRA	20.1	55.4*	43.8*	46.0*	52.5*	54.0*
AggPV	27.0	212.6*	90.1*	143.5*	184.3*	200.6*
Employer						
ERC	2.3	3647.7*	208.0*	546.8*	2713.9*	4477.8*
ERCPerc5%	0.0	−10116.4*	−2011.6*	−2389.6*	−7039.4*	−10118.8*

Figure 3.19: Efficient Strategy Frontier - Low Scenario

The following graphs show the simulation results of the low interest rate scenario setup for a minimum coverage ratio of 90% (top row) respectively 80% (bottom row) for two different employee preference specifications. The vertical axis shows the aggregate employee preferences measured by the CRRA and PV specification respectively (from left to right). The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

90% minimum coverage ratio



80% minimum coverage ratio

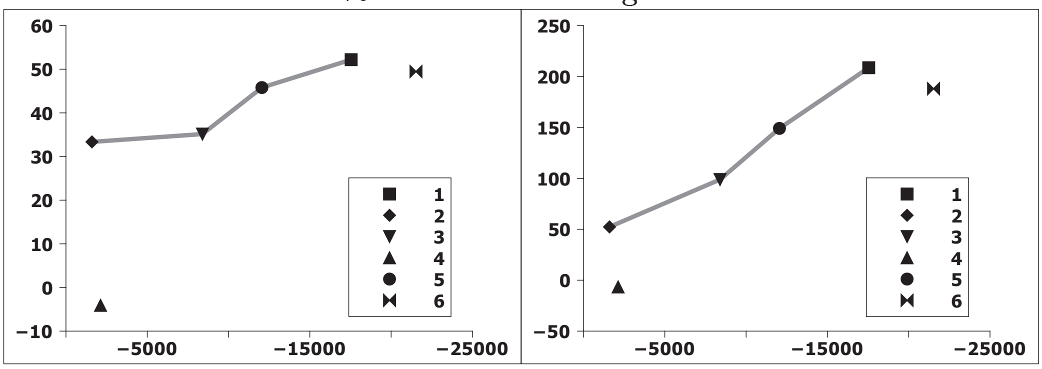
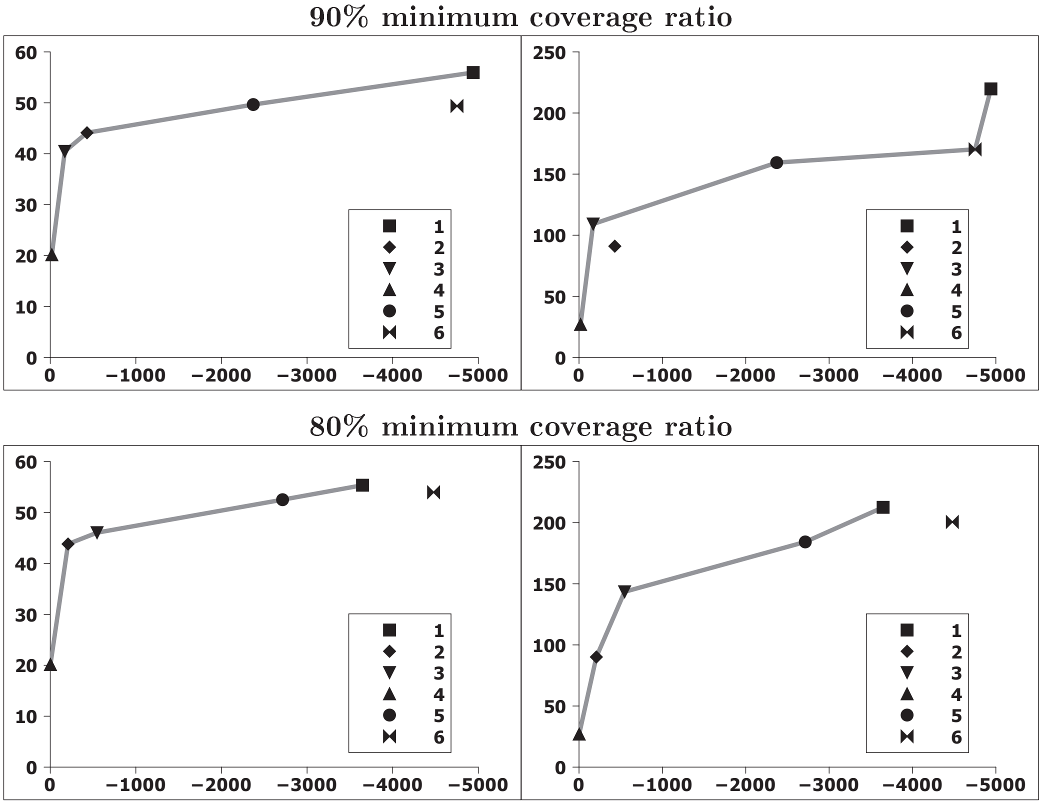


Figure 3.20: Efficient Strategy Frontier - High Scenario

The following graphs show the simulation results of the high interest rate scenario setup for a minimum coverage ratio of 90% (top row) respectively 80% (bottom row) for two different employee preference specifications. The vertical axis shows the aggregate employee preferences measured by the CRRA and PV specification respectively (from left to right). The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.



The results show that when a low rate scenario is assumed, the pension fund with the 80% minimum coverage ratio should choose one of the following strategies: 50/50, CPPI, PORS or FEQ. The choice depends on the risk capacity of the employer. For this pension fund and the given environment the two strategy options CASH and SLS are not efficient, as they are dominated independent of the employees preference representation. For a 90% minimum coverage ratio fund, the choice of efficient strategies is even narrower. It should choose among the 50/50, PORS and FEQ strategy under CRRA employee preferences and among 50/50, PORS, SLS and FEQ under PV preferences. (see top row in figure 3.19). For the high rate scenario the situation is different. If the minimum coverage ratio is set at 80%, all strategies except the SLS strategy are efficient. It is however noteworthy, that the more aggressive strategies, PORS and FEQ do not offer considerably more utility to employees compared to the additional risk they entail for the employer independent of the preference specification. Choosing between 50/50 and CPPI may therefore seem reasonable. For a 90% coverage ratio setup the situation is similar, however, when it is assumed that the preferences of the aggregated employees are represented by prospect theory, the CPPI strategy should be preferred over the 50/50 approach as the dynamic strategy dominates the semi-passive balanced approach.

In order to see how the relative attractiveness of the CPPI versus the 50/50 strategy depends on the choice of the fixed rate, the dominance limits for the low and high interest rate scenarios are shown in figures 3.21 and 3.22. As is depicted, the break even-rate for the employee preferences is shifted to the left in the low interest rate scenario while except for the aggregated CRRA preference measure the break-even rates are higher in the high interest rate scenario when compared to the base case scenario. (Table 3.5 in the appendix gives an overview of the break even rates for all considered scenarios.) These results show that the relationship between the fixed rate and the assumed long term risk free rate is crucial for the attractiveness of the CPPI strategy. If it is assumed that the risk free rate will be very low in the long run, choosing a dynamic CPPI strategy over a 50/50 strategy would only make sense if the fixed rate is set at a very low level between 0% and 0.8%, depending on whether the employee preferences are represented by a CRRA utility function or a PV function. However, if it is assumed that the interest rate will be high in the long run, a fixed rate of as high as 3% may be chosen if the preferences of the employees are modeled with a PV function. Thus, the choice of an efficient strategy may depend crucially on the assumptions about how the broad market and in particular the risk free interest rate will behave in the long run.

Figure 3.21: Dominance Limits - Low Scenario
The following graphs show the aggregate employee preferences (top), the employee voting preferences (middle) as well as the employer preferences (bottom) as a function of the fixed rate for the 50/50 as well as the CPPI strategy in the low interest rate scenario. The aggregate employee preferences correspond to the CRRA and PV specifications as defined in equations 3.14 to 3.16. The employee voting preferences depict the percentage of cohorts for which the specified strategy is preferred given the fixed rate. The employer preferences correspond to the ERC and ERCPerc risk measures as defined in equations 3.17 to 3.18.

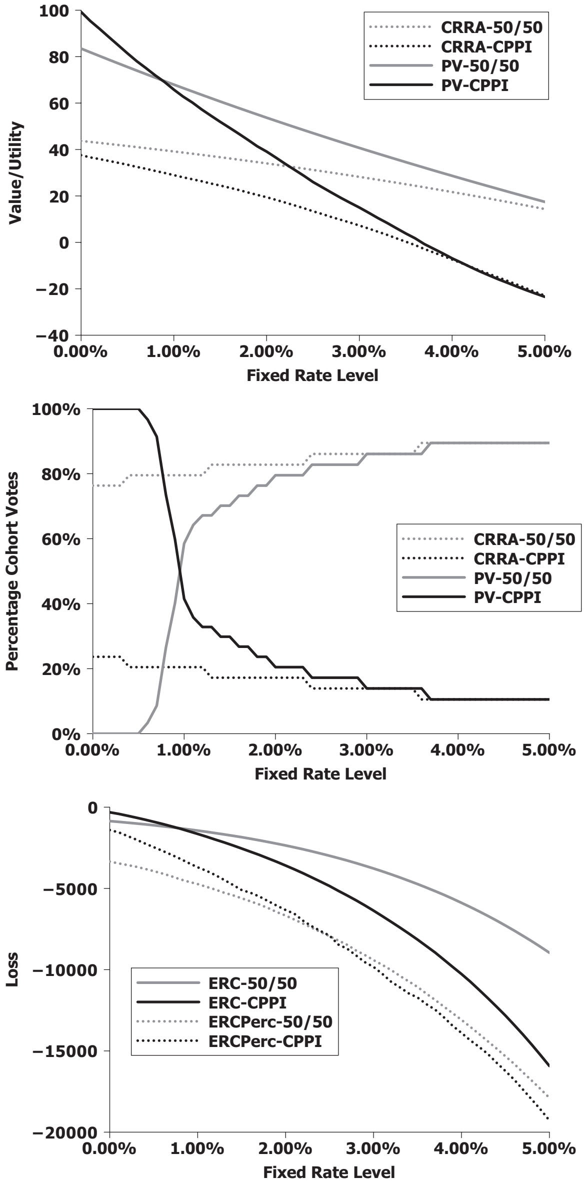
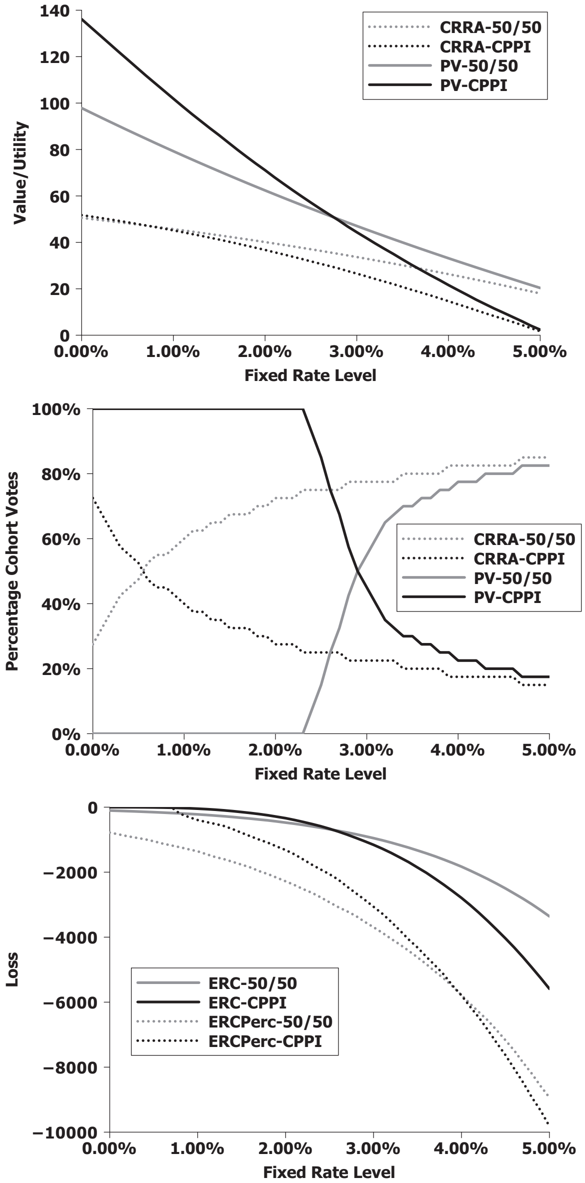


Figure 3.22: Dominance Limits - High Scenario

The following graphs show the aggregate employee preferences (top), the employee voting preferences (middle) as well as the employer preferences (bottom) as a function of the fixed rate for the 50/50 as well as the CPPI strategy in the high interest scenario. The aggregate employee preferences correspond to the CRRA and PV specifications as defined in equations 3.14 to 3.16. The employee voting preferences depict the percentage of cohorts for which the specified strategy is preferred given the fixed rate. The employer preferences correspond to the ERC and ERCPerc risk measures as defined in equations 3.17 to 3.18.



3.4.4 Block-Bootstrap Scenario

The simulation results obtained so far have been generated by using a Monte Carlo simulation framework where the risky asset is modeled by a Geometric Brownian motion and the interest rate is retrieved from a CIR model. This approach has the advantage of being able to capture well the broad features of the market's risk and return characteristics while being parsimonious with respect to the number of parameters which need to be defined. However, some features of the actual market behavior is not captured by the applied Monte Carlo simulation approach such as the heteroskedasticity and possible auto-correlation effects which are observed in actual market data series. In order to assess the possible impact of using a simplified simulation framework, in this subsection the base case analysis of subsection 3.4.1 is replicated, whereby instead of a Monte Carlo simulation a block-bootstrap simulation approach is used to generate the return series. The block-bootstrap approach applied to generate the return series corresponds closely to the one applied by Annaert, Osselaer and Verstraete (2009). The used data sets are the daily US equity market and 1-month US treasury series used by Fama and French (1993) for the time period of July 1926 until September 2013. In order to generate one-year of data for the simulation, the block-bootstrap approach draws randomly a one-year block from the actual data series[38]. Thereby, each one-year block consists of 252 observations. The reason for drawing one-year blocks of data instead of daily observations is that by drawing blocks the auto-correlation and heteroskedasticity properties of the actual data series are preserved. For the next year, again a random date is drawn from the entire data set and the corresponding one-year block of data is used as the second year's return. This procedure is repeated 40 times in order to generate the 40-year return series which resembles one simulation path used in this study. Similar to the Monte Carlo approach, 5000 paths are generated in this way. The fixed rate is set at 2% and a setup with a 90% minimum coverage ratio as well as a 80% minimum coverage ratio are considered. The corresponding results are listed in table 3.4 and the corresponding efficient strategy frontiers are depicted in figure 3.23.

[38]The equity market and interest rate returns are drawn together, thus the combination of equity market returns and interest rate returns correspond to actual observations.

Table 3.4: Block-Bootstrap Scenario

The table shows the results of a block-bootstrap simulation with 5000 runs. Two minimum coverage ratio settings are considered (90% and 80%). For each coverage ratio setting and investment strategy the risk return metrics, as well as the employee and employer preferences as specified in subsection 3.2.2 are obtained. In each row, except for the employer preferences, the * and ° symbols indicate that the mean is significantly higher compared to the next lower mean at the 1% respectively 5% significance level. For the employer risk measures the * and ° symbols indicate that the mean is significantly lower compared to the next higher mean at the 1% respectively 5% significance level. Thereby, a two-sided t-test for the difference in means with unknown variances is conducted in every line except for the ERCPerc5% risk measures where a bootstrap approach is used to determine the significance.

	CASH	**FEQ**	**50/50**	**CPPI**	**PORS**	**SLS**
Min Coverage Ratio 90%						
Risk Return Metrics						
Mean Annual Compounded Return	3.3%	9.2%*	6.8%*	7.3%*	8.7%*	8.1%*
Mean Annual Standard Deviation	3.0%	19.6%*	9.8%*	13.2%*	17.1%*	17.4%*
Mean Sharpe Ratio	0.00	0.32	0.36*	0.29*	0.32*	0.28*
Employee						
AggCRRA	12.9	60.6*	45.7*	46.6*	57.2*	54.7*
AggPV	11.7	245.3*	79.5*	117.2*	193.2*	179.7*
Employer						
ERC	−6.8	−12539.4*	−1106.9*	−978.8*	−8392.4*	−9810.8*
ERCPerc5%	0.0	−29721.8*	−4036.6*	−2621.5*	−16974.6*	−18439.6*
Min Coverage Ratio 80%						
Risk Return Metrics						
Mean Annual Compounded Return	3.3%	9.2%*	6.8%*	7.8%*	8.8%	8.7%*
Mean Annual Standard Deviation	3.0%	19.6%*	9.8%*	15.6%*	18.4%*	18.9%*
Mean Sharpe Ratio	0.00	0.32*	0.36*	0.28*	0.31*	0.29*
Employee						
AggCRRA	12.9	60.1*	45.3*	50.4*	58.3*	58.4
AggPV	11.7	237.8*	78.3*	146.8*	210.7*	217.0
Employer						
ERC	−0.2	−10402.3*	−639.4*	−2911.3*	−9977.0*	−12526.1*
ERCPerc5%	0.0	−26764.0*	−3390.1*	−8166.3*	−22926.2*	−26681.1*

The results using the block-bootstrap approach are qualitatively similar to the base case results applying the Monte Carlo simulation approach as listed in table 3.1. However, there are some key differences. With regards to the risk metrics, generally the return as well as the standard deviations of the different strategies are higher with the exception of the compounded CASH return which is lower. Also, the Sharpe ratios are higher in the bootstrap setup compared to the base case. An interesting observation is made with respect to the employer contribution, which are considerably higher independent of the risk metrics and strategies compared to the base case. Further, the CPPI strategy now exhibits slightly higher compounded returns compared to the 50/50 strategy even under the 90% minimum coverage ratio setup.

Moving to the relative attractiveness of the strategies from an employer's and employee's point of view, two changes are observed compared to the base case with regards to the 90% minimum coverage ratio setup. First, the SLS approach is now dominated under both employee preference specifications which was not the case in the base case setup. Further, the 50/50 approach is now dominated even under the CRRA employee preference specification. Thus, with regards to the crucial and extensively discussed relative attractiveness of the 50/50 and CPPI strategies, the later is becoming more attractive relative to the former at least with respect to the employee's preferences. This finding is confirmed by the dominance limits which are depicted in figure 3.24. The break-even rates from an employee's perspective under aggregated CRRA and PV preferences are at 2.3% respectively 4.2% in the block bootstrap scenario. Further, with regards to the voting preferences, the break even levels are at 2.1% respectively 4.2%. Thus, all break even rates from an employee's point of view are higher compared to the base case. However, moving to the employer's view it is seen that for the ERC preferences, the break-even rate is at 2.3% which is lower than the 2.6% in the base case. On the other side, based on the ERCPerc measure the break-even rate is noticeably higher at 5% vs. 3.9% in the base case.

Figure 3.23: Efficient Strategy Frontier - Block-Bootstrap Scenario
The following graphs show the simulation results of the block-bootstrap scenario setup for a minimum coverage ratio of 90% (top row) respectively 80% (bottom row) for two different employee preference specifications. The vertical axis shows the aggregate employee preferences measured by the CRRA and PV specification respectively (from left to right). The horizontal axis represents the expected employer contribution as defined in equation 3.17. The different strategies are numbered as follows: 1=FEQ, 2=50/50, 3=CPPI, 4=CASH, 5=PORS, 6=SLS. The strategies which are not dominated are connected by the line.

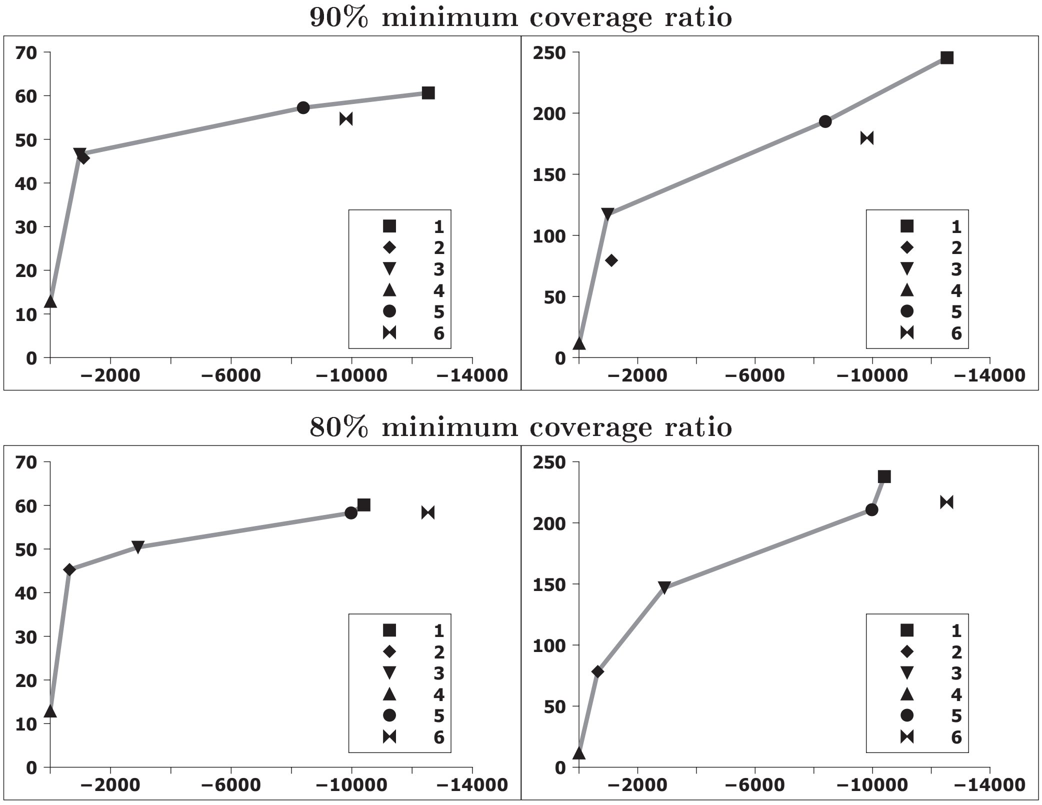
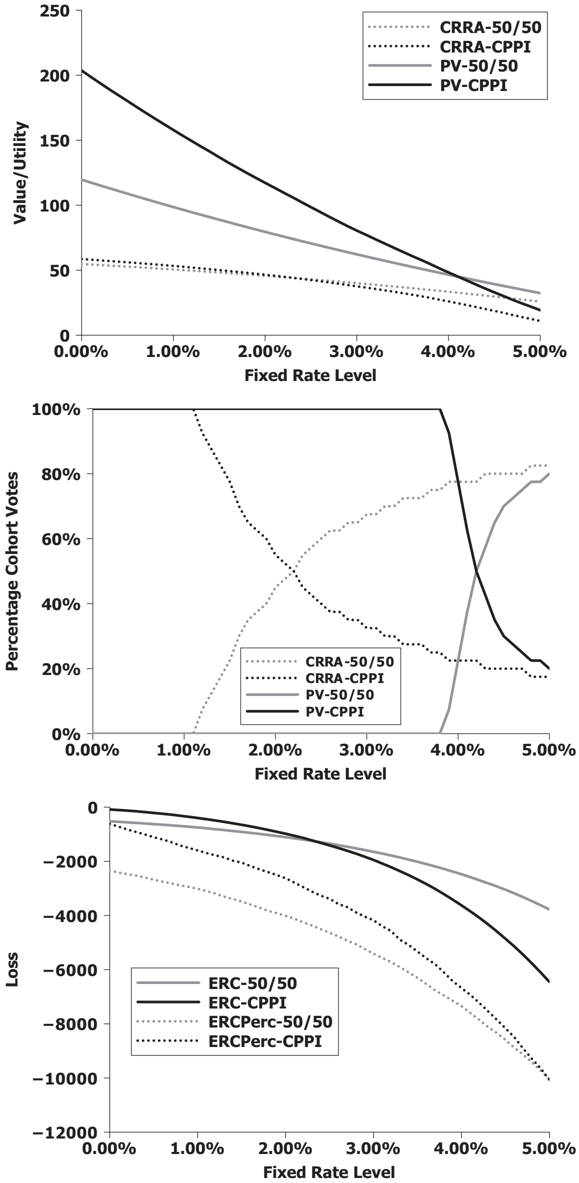


Figure 3.24: Dominance Limits - Block-Bootstrap Scenario

The following graphs show the aggregate employee preferences (top), the employee voting preferences (middle) as well as the employer preferences (bottom) as a function of the fixed rate for the 50/50 as well as the CPPI strategy in the block-bootstrap scenario. The aggregate employee preferences correspond to the CRRA and PV specifications as defined in equations 3.14 to 3.16. The employee voting preferences depict the percentage of cohorts for which the specified strategy is preferred given the fixed rate. The employer preferences correspond to the ERC and ERCPerc risk measures as defined in equations 3.17 to 3.18.



3.5 Conclusion

This study introduces a pension fund model simulation framework of the Swiss pension system with the aim to test whether dynamic portfolio insurance strategies are worthwhile of being considered as alternatives to buy-and-hold and semi-dynamic strategies in this framework. The three dynamic strategies under consideration have been analyzed and compared in previous studies[39]. A key feature of the pension fund framework is the peculiarity that the preferences of the employees (investors) as well as the employer are modeled, as the employer is liable at least in part for any restructuring costs in the given setting. The preferences of the employees are modeled by using a classic expected utility function (CRRA) as well as using a prospect theory value (PV) function. In order to determine the relative attractiveness of the various strategies in the given framework, a Monte Carlo simulation is conducted. Thereby, in the base case scenario the parameters are chosen based on the choices of previous studies. The results of the base case study show that to a large extent the preferences for the strategies depend on the risk return properties of the strategies. In general, the more risky strategies are favored by the employees and the less risky strategies are favored by the employer. However, it was shown that certain strategies may be dominated by others, meaning that there exists another strategy which is preferred from the employee's as well as the employer's point of view. In the base case it was shown that under a 80% minimum coverage ratio constraint the SLS is dominated by the FEQ strategy. Further, when the preferences of the employees are modeled under prospect theory value preferences the CPPI strategy dominates a semi-dynamic 50/50 strategy. This result is of particular interest as the 50/50 strategy can be seen as a proxy of strategies which are typically applied in practice by pension funds (see figure 3.3).

In order to corroborate the results of the base case, a sensitivity analysis is conducted in section 3.4.2. The results indicate that the choice of the fixed rate has a key impact on the relative preferences for the different strategies under consideration. It is shown that the higher the fixed rate is set, the less attractive become the portfolio insurance strategies CPPI and SLS. With

[39]See for example Annaert, Osselaer and Verstraete (2009); Dierkes, Erner and Zeisberger (2010); Dichtl and Drobetz (2011).

regards to the relative attractiveness of the CPPI and 50/50 strategy the dominance limits as a function of the fixed rate were established in order to determine at which level the CPPI strategy would no longer dominate the 50/50 strategy. As was found in the analysis, depending on the choice of the preference representation of the employer and the employee, the break even fixed rate lies between 0.5% and 2.8%[40]. The sensitivity analysis is extended in a second step in order to investigate the impact of different demographic weighting schemes. It is shown that in general the weighting scheme does not have a large impact with regards to the relative attractiveness of the strategies under consideration. However, with regards to the break-even rate between the CPPI and 50/50 strategy it is found that an over-weighting of the older cohorts leads to slightly lower break-even rates while an over-weighting of the younger cohorts modestly increases them[41]. However, overall the sensitivity results show that the choice of the fixed rate is more crucial in determining the relative attractiveness of the strategies.

In order to show how the assumption about the market parameters influence the results, a case study was conducted in subsection 3.4.3. Thereby, two extreme scenarios based on actual historical time periods are chosen to determine the simulation parameters. The low scenario represents a low interest rate scenario similar to the time period between 1930 and 1969 where the average risk free rate stood at 1.65%. The other scenario represented a high interest rate scenario where it is expected that the long term average risk free rate stands at 5.6% as during the period between 1970 and 2009. As the results show, varying the interest rate while keeping the fixed rate steady, has a similar effect as does the varying of the fixed rate but in the opposite direction. With a low average interest rate the CPPI strategy becomes less attractive while with a high average interest rate it becomes more attractive relative to the other strategies. The bottom line results show that the

[40]See table 3.5 in the appendix for an overview of the dominance limits under all considered scenario cases.

[41]The reason for this result is related to the redistribution effects which are present in the applied pension framework. Further insights regarding this topic are given in the appendix.

attractiveness of the CPPI strategy relative to the other strategies largely depends on the difference between the assumed long term risk free rate and the fixed rate. The smaller the fixed rate relative to the risk free rate the more attractive becomes the CPPI strategy and vice versa.

The last section investigated whether the use of actual return data via a block-bootstrap approach has any impact on the overall findings. The results reveal that under the base case scenario the SLS is dominated in all considered preference specifications[42]. Also, the CPPI approach gains attractiveness relative to the 50/50 approach. This is reflected in the dominance break even rates for the block-bootstrap case which lie between 2.1% and 4.1%, depending on the choice of the preference measures (see also table 3.5 in the appendix).

This study is the first to the best of the author's knowledge, to investigate the attractiveness of portfolio insurance strategies in the context of a pension system where the preferences of the employees as well as the employer are jointly considered. The most significant result is found with respect to the relative attractiveness of the CPPI strategy and the 50/50 strategy. As it was shown, there may exist constellations where a dynamic insurance strategy in the form of a CPPI approach is preferred over a 50/50 strategy from the employer's as well as the employee's point of view. However, the study also revealed that this result crucially depends on the way the preferences of the employee as well as the employer are modeled. Further, the difference between the assumed long term risk free rate and the fixed rate is key for the relative attractiveness of the strategies. With respect to the other two dynamic strategies, the results of this study indicate that the PORS approach can be seen as a less risky version of the full equity investment strategy FEQ from the employer's point of view. Further, the results find that the SLS approach is not particularly attractive in the given setting as for a large variety of simulation setups considered in this study the SLS approach was

[42] Assuming a fixed rate of 2% as in the base case.

dominated by the FEQ strategy.

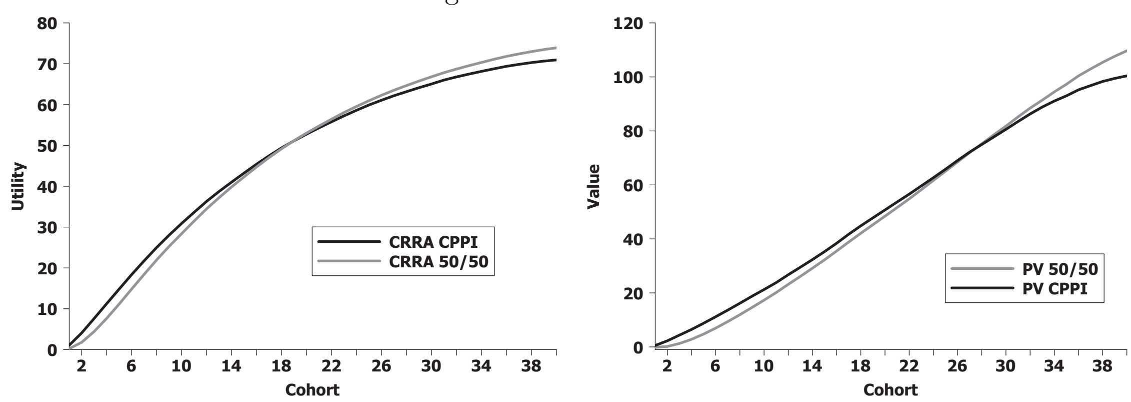
Future research in this field could focus on mapping the pension fund investor's preferences more precisely, possibly based on a survey. Further, the incorporation of more advanced dynamic insurance strategies in a similar framework, such as variable proportion portfolio insurance strategies (VPPI) as analyzed for example in Zieling, Mahayni and Balder (2014), may yield additional valuable insights with respect to the attractiveness of portfolio insurance strategies in the context of a pension fund framework. Also, the investigation of strategies which protect accrued capital gains, such as the time invariant portfolio protection (TIPP) strategy as first proposed by Estep and Kritzman (1988), may be an interesting starting point for further research in the given framework.

3.6 Appendix

Cohort Contribution at Break Even Rate

Figure 3.25: Per Cohort Contribution (CPPI versus 50/50) - Base Case Scenario at Break Even Rate

The graphs show the average CRRA utility $\frac{1}{M}\sum_{m=1}^{M} CRRA(\Delta x_{y,m})$ (left graph) respectively the average prospect value $\frac{1}{M}\sum_{m=1}^{M} v(\Delta x_{y,m})$ (right graph) of every cohort y for the 50/50 as well as the CPPI strategy for the equal-demographic weighting scheme under the 90% minimum coverage ratio setup. For the analysis a 2.7% fixed rate is used. The cohorts are ordered along the horizontal axis from the oldest to the youngest cohort. The utility respectively the prospect value of the cohorts for the two strategies is referenced on the vertical axis on the left hand side.



Sensitivity Analysis

This section of the appendix gives detailed insights into the sensitivity analysis conducted in subsection 3.4.2. In particular it will be shown how the demographic weighting specification has an impact on the dominance relationship between the 50/50 and CPPI strategy. For the sake of brevity the analysis will be conducted only with respect to the PV preference specification, noting that the same conclusions would be drawn using the CRRA specification. Also, the illustrations made in this subsection refer to the base case scenario and a 90% minimum coverage ratio unless stated otherwise.

As it was shown in the sensitivity analysis in section 3.4.2, changing the demographic weighting impacts the dominance limits of the CPPI and 50/50 strategy in such a way that the aggregated preferences of the employees for the 50/50 relative to the CPPI strategy increase when the old cohorts are over-weighted. The opposite is the case when the young cohorts are over-weighted. Figure 3.26 illustrates the preference contribution in the old and young-demographic weighting schemes similar to the bottom right graph in figure 3.6 for the equal weighted scheme. Compared to the equal-demographic weighting scheme (bottom right graph in figure 3.6), the older cohorts contribute considerably more to the average aggregated preferences in the old-demographic weighting scheme. The same is true with respect to the younger cohorts in the young-demographic weighting scheme. This is not surprising as the old respectively the young cohorts are overrepresented in the two alternative weighting schemes. What is hidden in figure 3.26, is the fact, that changing the weighting scheme changes the distribution of wealth among the cohorts when comparing different strategies. This is seen in figure 3.27, where the relative contributions per cohort in the old and young-demographic weighting schemes are shown net of the over-weighting effect. If a change in the weighting scheme did not have an effect on the relative wealth of the different cohorts, the graphs in figure 3.27 would be the same as the bottom right one in figure 3.6.

Figure 3.26: Relative Value Contribution per Cohort - Old and Young Demographic Weighting
The two graphs visualize the PV preference contribution of each cohort with respect to the two strategies (CPPI&50/50) for the old-demographic weighting scheme (left) and the young-demographic weighting scheme (right). The bars represent the difference between the PV value of the CPPI strategy minus the PV value of the 50/50 strategy for a specific cohort in a scenario with a 2% fixed rate. The PV for each cohort and strategy is determined as follows whereby the weighting factor d_y is applied: $\frac{1}{M}\sum_{m=1}^{M} d_y \cdot v(\Delta x_{y,m})$.

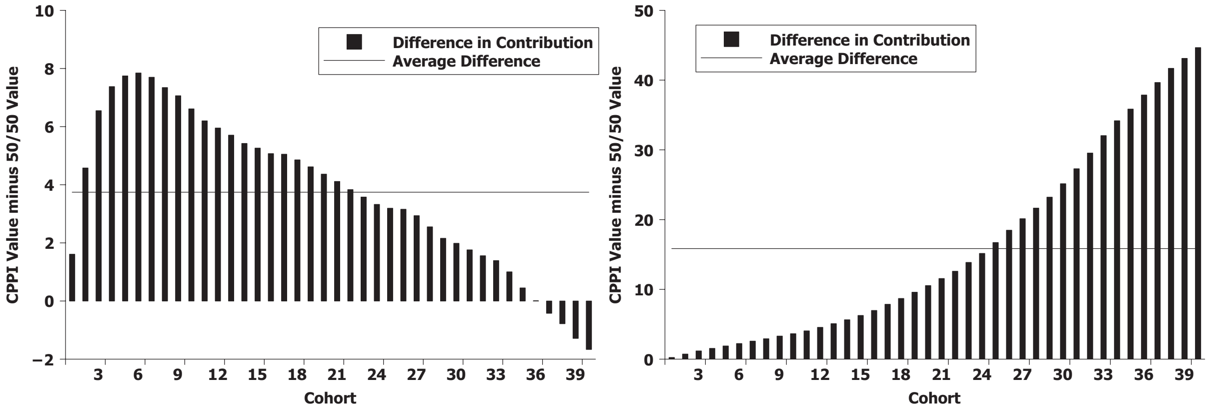
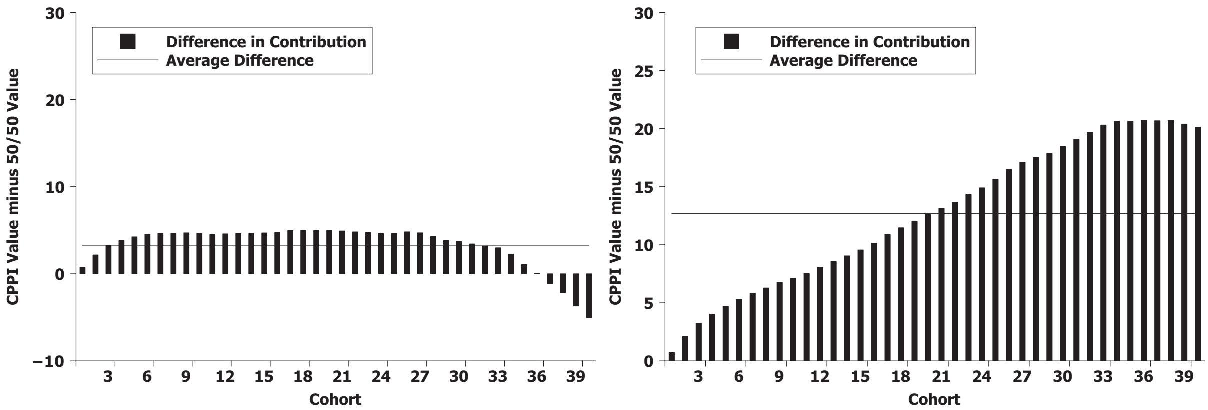


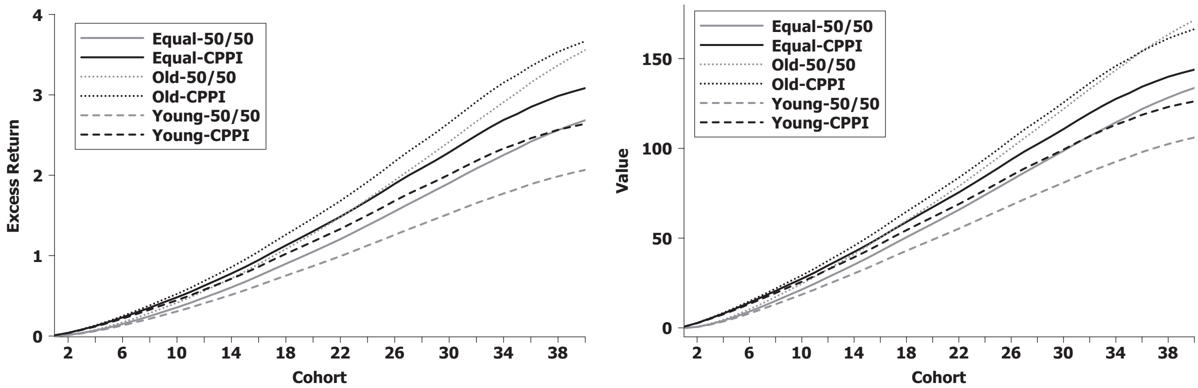
Figure 3.27: Relative Value Contribution per Cohort - Net of Weighting Effect
The two graphs visualize the PV preference contribution of each cohort with respect to the two strategies (CPPI&50/50) for the old-demographic weighting scheme (left) and the young-demographic weighting scheme (right). Thereby, the contributions net of the over-weighting effect are shown in this figure. The bars represent the difference between the average PV of the CPPI strategy minus the average PV of the 50/50 strategy for a specific cohort in a scenario with a 2% fixed rate. The PV for each cohort and strategy is determined as follows whereby no weighting factor is used: $\frac{1}{M}\sum_{m=1}^{M} v(\Delta x_{y,m})$.



The encountered redistribution of wealth can be explained by a positive cushion effect which is much more pronounced under the old demographic weighting scheme and less pronounced under the young demographic weighting scheme. The positive cushion effect stems from reserves which are accumulated when cohorts retire while the coverage ratio of the fund is larger than 100%. For example, when the coverage ratio stands at 110% and a cohort retires, this cohort only receives a payout of 100% of the liabilities, although assets for covering liabilities of 110% would be available. The difference goes into the reserves which increases the coverage ratio and benefits the cohorts which are still active. Thus when large cohorts retire early, as it is the case in the old weighting scheme, an additional positive cushion can be accumulated for the following cohorts. There is also a negative cushion effect, which may occur when the coverage ratio is below 100% but above the minimum required coverage ratio. When a large cohort retires in this situation, the cohort receives a payout of 100% while the difference to the actual coverage ratio is imposed on the remaining cohorts. This negative cushion effect is also more pronounced if the old cohorts are over-weighted. When aggregating the two effects, the positive cushion effect dominates which leads to higher expected payouts under the old-demographic weighting scheme for for all cohorts except for the one next to retirement. The opposite is true when the young-demographic weighting scheme is used. These effects are illustrated in figure 3.28, where the left graph depicts the average excess return under the different weighing schemes and the right graph shows the corresponding PV preferences.

Figure 3.28: Prospect Value and Excess Return under Different Weighting Schemes

The graphs show the average excess return $\frac{1}{M}\sum_{m=1}^{M}\Delta x_{y,m}$ (left graph) respectively the average prospect value $\frac{1}{M}\sum_{m=1}^{M}v(\Delta x_{y,m})$ (right graph) of every cohort y for the 50/50 as well as the CPPI strategy for the equal-demographic, old-demographic and young-demographic weighting schemes. For the analysis a 2% fixed rate is used. The cohorts are ordered along the horizontal axis from the oldest to the youngest cohort. The excess return respectively the prospect value of the cohorts for the two strategies is referenced on the vertical axis on the left hand side.



Higher or lower returns per se do not explain why the relative attractiveness of the CPPI and 50/50 strategy changes from an employee's point of view when changing the weighting scheme. However, when analyzing the right graph of figure 3.28, it can be seen that for the youngest cohorts under the old demographic weighting scheme, the 50/50 strategy is preferred over the CPPI strategy despite the later providing them with a higher mean excess return. This finding indicates that the disadvantageous distributional properties of the CPPI strategy relative to the 50/50 strategy, which was already addressed in the the base case analysis in section 3.4, is the main reason of why different demographic weighting schemes influence the relative attractiveness of the two strategies under consideration. This point is illustrated more clearly in figure 3.29, where for every cohort the difference in the value contribution as depicted in figure 3.27 is broken down into the contribution from positive returns (left graph) and the contribution from negative returns (right graph). As it is seen, the contribution from the positive returns actually become negative for the youngest cohorts under the old-demographic weighting scheme, although the mean expected excess return of the CPPI strategy for these cohorts is higher compared to the 50/50 strategy. On the other side, the positive return contributions are all positive and increasing when moving from the old cohorts to the young cohorts under the young-demographic weighting scheme. When looking at the contributions from the negative returns (right graph in figure 3.29) it is interesting to see that these contributions also deteriorate when moving from a young-demographic weighting scheme to an old demographic weighting scheme. The reason for this also has to do with the distributional properties of the CPPI strategy and the way the restructuring mechanism is set up in the pension fund framework. The graph in figure 3.30 depicts the average negative return for each cohort under each weighting scheme for the CPPI strategy as well as the 50/50 strategy[43]. As shown, the trough of the lowest average returns are shifted to the right towards younger cohorts for the CPPI strategy compared to the 50/50 strategy. This effect is due to the fact that the under the CPPI strategy a restructuring is more likely to be avoided in the early years, when the old cohorts (cohorts on the left side of the graph) retire. However, the avoidance of a restructuring still results in a lower coverage ratio which in turn increases the floor used in the CPPI strategy and thereby reduces the return potential. As a result, the restructuring probability for the following years is increased whereby the trough of the incurred negative returns is shifted to younger cohorts relative to the 50/50 strategy as shown in figure 3.30. While the restructuring costs under a CPPI strategy are lower, they are more sensitive to alternative demographic weighting schemes. If old cohorts are over-weighted, the potential restructuring costs for the younger cohorts are increased via the negative cushion effect. This is illustrated by the dotted line in figure 3.29. The 50/50 strategy is much less sensitive to changes to the demographic weighting scheme as early restructurings are not avoided and the oldest cohorts contribute their share to the restructuring costs (see grey lines in figure 3.30). Figure 3.30 in combination with the left graph

[43] For each cohort the numbers are obtained by filtering out the negative returns observed from all simulation paths in the base case and then calculating the average negative return.

in figure 3.26 also indicates why from an employer's perspective when using the ERCPerc measure, the CPPI strategy is always preferred under the old-demographic weighting scheme. When using a 50/50 strategy the largest average negative returns relative to the CPPI strategy occur when the oldest cohorts retire which are the cohorts which are heavily overrepresented in the old-demographic weighting scheme. On the other side, the cohorts for which the 50/50 strategy generates less severe average negative returns relative to the CPPI strategy are under-weighted in this specific weighting scheme.

Figure 3.29: Relative Value Contribution per Cohort - Breakdown of Positive and Negative Return Contributions
The graphs show the break down of positive and negative return contributions of the difference between the average PV of the CPPI strategy minus the average PV of the 50/50 strategy as they are shown in figure 3.27 for the different weighting schemes (equal, old and young). The graph on the left side shows the contribution of the positive returns to the out/under-performance of the CPPI strategy relative to the 50/50 strategy. The graph on the right side shows the contribution of the negative returns to the out/under-performance of the CPPI strategy relative to the 50/50 strategy.

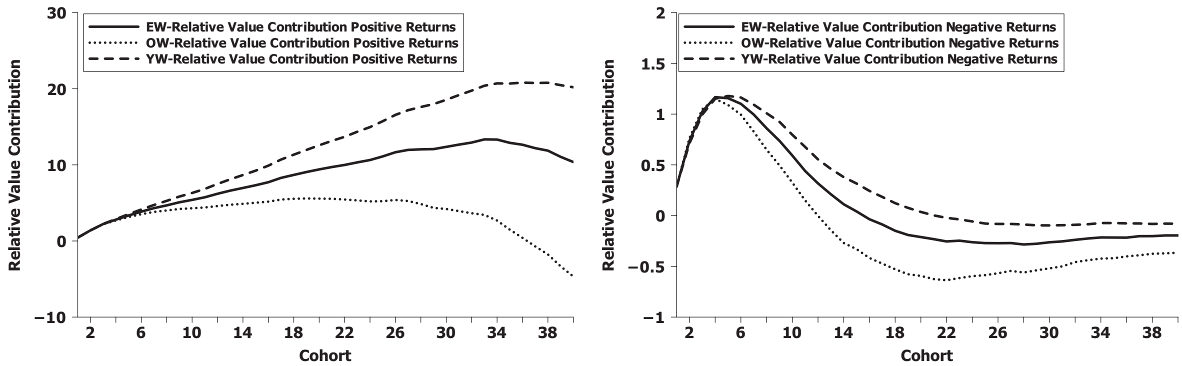
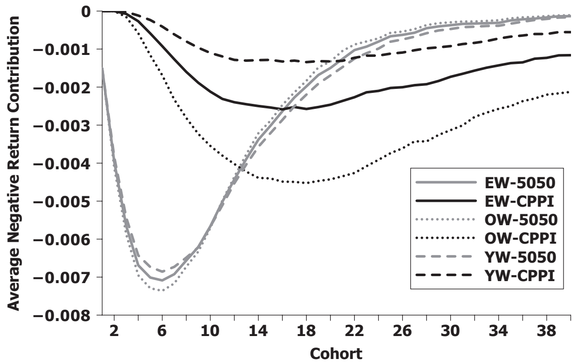


Figure 3.30: Average of Negative Excess Returns per Cohort under Different Weighting Schemes

This graph shows the average of the negative excess returns $\frac{1}{M^-}\sum_{m=1}^{M^-}(\Delta x_{y,m})$ for the CPPI and 50/50 strategy under the different weighting schemes (equal, old, young). M^- refers to the number of simulation outcomes for which a negative excess return is observed. The cohorts are ordered along the horizontal axis from the oldest to the youngest cohort. The excess return of the cohorts for the two strategies is referenced on the vertical axis on the left hand side.



Dominance Limits

Table 3.5: Overview of Dominance Limits (CPPI versus 50/50)

This table lists the break even rates from the employee's as well as the employer's point of view for the six scenarios considered in this study. The break-even rate refers to the rate at which the employee respectively the employer is indifferent between preferring the CPPI strategy over the 50/50 strategy. The break-even rates depend on the choice of the risk measure. From the employee's view the break even rate is listed for the following risk measures: Aggregated CRRA utility (AggCRRA), cohort votes under CRRA (VoteCRRA), aggregated prospect theory values (AggPV) and cohort votes under PV (VotePV). From the employer's view the expected restructuring contributions (ERC) as well as the expected restructuring contribution percentile (ERCPerc) measures are shown. The L and the H, indicate the combination of employer and employee preferences for which the dominance rate is the lowest respectively the highest in a given scenario.

		AggCRRA	VoteCRRA	AggPV	VotePV
	Base Case	0.7%	0.5%	2.7%	2.8%
ERC	2.5%		L		
ERC Perc	3.9%				H

		AggCRRA	VoteCRRA	AggPV	VotePV
	Old Case	0.4%	0.7%	2.4%	2.9%
ERC	2.5%	L			
ERC Perc	$> 5\%$				H

		AggCRRA	VoteCRRA	AggPV	VotePV
	Young Case	1.3%	1.1%	3.1%	3.1%
ERC	2.7%		L		
ERC Perc	3.7%			H	

		AggCRRA	VoteCRRA	AggPV	VotePV
	Low Case	$< 0\%$	$< 0\%$	0.8%	0.9%
ERC	2.5%	L	L		
ERC Perc	3.9%				H

		AggCRRA	VoteCRRA	AggPV	VotePV
	High Case	0.7%	0.4%	3.0%	3.0%
ERC	3.2%		L		
ERC Perc	$> 5\%$				H

		AggCRRA	VoteCRRA	AggPV	VotePV
	Boot. Case	2.3%	2.1%	4.1%	4.1%
ERC	2.3%		L		
ERC Perc	5.0%				H

Bibliography

Abdellaoui, Mohammed. 2000. "Parameter-free elicitation of utility and probability weighting functions." *Management Science*, 46(11): 1497–1512.

Agarwal, V., and N.Y. Naik. 2004. "Risks and portfolio decisions involving hedge funds." *Review of Financial Studies*, 17(1): 63–98.

Alberg, Dima, Haim Shalit, and Rami Yosef. 2008. "Estimating stock market volatility using asymmetric GARCH models." *Applied Financial Economics*, 18(15): 1201–1208.

Andrews, D.W.K., I. Lee, and W. Ploberger. 1996. "Optimal change-point tests for normal linear regression." *Journal of Econometrics*, 70(1): 9–38.

Angelidis, Timotheos, Alexandros Benos, and Stavros Degiannakis. 2007. "A robust VaR model under different time periods and weighting schemes." *Review of Quantitative Finance and Accounting*, 28(2): 187–201.

Annaert, Jan, Sofieke Van Osselaer, and Bert Verstraete. 2009. "Performance evaluation of portfolio insurance strategies using stochastic dominance criteria." *Journal of Banking & Finance*, 33(2): 272–280.

Arnott, Robert D, and Peter L Bernstein. 2002. "What Risk Premium Is "Normal"?" *Financial Analysts Journal*, 58(2): 64–85.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. "Coherent measures of risk." *Mathematical finance*, 9(3): 203–228.

Asness, Clifford S, Tobias J Moskowitz, and Lasse Heje Pedersen. 2013. "Value and momentum everywhere." *The Journal of Finance*, 68(3): 929–985.

Bai, J., and P. Perron. 2003. "Computation and analysis of multiple structural change models." *Journal of Applied Econometrics*, 18(1): 1–22.

Bali, Turan G, Stephen J Brown, and Mustafa Onur Caglayan. 2011. "Do hedge funds' exposures to risk factors predict their future returns?" *Journal of Financial Economics*, 101(1): 36–68.

Banz, R.W. 1981. "The relationship between return and market value of common stocks." *Journal of financial economics*, 9(1): 3–18.

Barone-Adesi, Giovanni, Kostas Giannopoulos, and Les Vosper. 2002. "Backtesting derivative portfolios with filtered historical simulation (FHS)." *European Financial Management*, 8(1): 31–58.

Bekiros, Stelios D, and Dimitris A Georgoutsos. 2005. "Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance." *Journal of International Financial Markets, Institutions and Money*, 15(3): 209–228.

Bender, J., R. Briand, F. Nielsen, and D. Stefek. 2010. "Portfolio of Risk Premia: A New Approach to Diversification." *The Journal of Portfolio Management*, 36(2): 17–25.

Black, Fischer, and AndreF Perold. 1992. "Theory of constant proportion portfolio insurance." *Journal of Economic Dynamics and Control*, 16(3): 403–426.

Black, Fischer, and Myron Scholes. 1973. "The pricing of options and corporate liabilities." *The journal of political economy*, 637–654.

Black, Fischer, and Robert C Jones. 1987. "Simplifying portfolio insurance." *The Journal of Portfolio Management*, 14(1): 48–51.

Black, Fischer, and Robert C Jones. 1988. "Simplifying portfolio insurance for corporate pension plans." *The journal of portfolio management*, 14(4): 33–37.

Bollen, N.P.B. 2011. "Zero-R2 Hedge Funds and Market Neutrality." *Workgin Paper. Vanderbuilt University.*

Bollen, N.P.B., and R.E. Whaley. 2009. "Hedge fund risk dynamics: Implications for performance appraisal." *The Journal of Finance*, 64(2): 985–1035.

Bollerslev, Tim. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics*, 31(3): 307–327.

Bollerslev, Tim. 1987. "A conditionally heteroskedastic time series model for speculative prices and rates of return." *The Review of Economics and Statistics*, 69(3): 542–547.

Bollerslev, Tim, and Jeffrey M Wooldridge. 1992. "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances." *Econometric Reviews*, 11(2): 143–172.

Broda, Simon A, Markus Haas, Jochen Krause, Marc S Paolella, and Sven C Steude. 2013. "Stable mixture GARCH models." *Journal of Econometrics*, 172(2): 292–306.

Brooks, Chris, AD Clare, JW Dalle Molle, and Gita Persand. 2005. "A comparison of extreme value theory approaches for determining value at risk." *Journal of Empirical Finance*, 12(2): 339–352.

Buraschi, A., R. Kosowski, and F. Trojani. 2011. "When there is no place to hide: Correlation Risk and the Cross-Section of Hedge Fund Returns." *Working Paper.*

Campbell, J.Y., and R.J. Shiller. 1991. "Yield spreads and interest rate movements: A bird's eye view." *The Review of Economic Studies*, 58(3): 495–514.

Carhart, M.M. 1997. "On persistence in mutual fund performance." *Journal of Finance*, 52(1): 57–82.

Christiansen, C., A. Ranaldo, and P. Söderlind. 2011. "The time-varying systematic risk of carry trade strategies." *Journal of Financial and Quantitative Analysis*, 46(4): 1107.

Christoffersen, Peter F. 1998. "Evaluating interval forecasts." *International economic review*, 39(4): 841–862.

Christoffersen, Peter F. 2012. *Elements of financial risk management.* Academic Press.

Cochrane, J.H., and M. Piazzesi. 2005. "Bond Risk Premia." *American Economic Review*, 138–160.

Cox, John C, Jonathan E Ingersoll Jr, and Stephen A Ross. 1985. "A theory of the term structure of interest rates." *Econometrica: Journal of the Econometric Society*, 53(2): 385–407.

Davidson, James. 2004. "Moment and memory properties of linear conditional heteroscedasticity models, and a new model." *Journal of Business & Economic Statistics*, 22(1): 16–29.

Detemple, Jérôme, and Marcel Rindisbacher. 2008. "Dynamic asset liability management with tolerance for limited shortfalls." *Insurance: Mathematics and Economics*, 43(3): 281–294.

Dichtl, Hubert, and Wolfgang Drobetz. 2011. "Portfolio insurance and prospect theory investors: Popularity and optimal design of capital protected financial products." *Journal of Banking & Finance*, 35(7): 1683–1697.

Dierkes, Maik, Carsten Erner, and Stefan Zeisberger. 2010. "Investment horizon and the attractiveness of investment strategies: A behavioral approach." *Journal of Banking & Finance*, 34(5): 1032–1046.

Di Giacinto, Marina, Salvatore Federico, and Fausto Gozzi. 2011. "Pension funds with a minimum guarantee: a stochastic control approach." *Finance and Stochastics*, 15(2): 297–342.

Dimitrakopoulos, Dimitris N, Manolis G Kavussanos, and Spyros I Spyrou. 2010. "Value at risk models for volatile emerging markets equity portfolios." *The Quarterly Review of Economics and Finance*, 50(4): 515–526.

Dupuis, DJ. 1999. "Exceedances over high thresholds: A guide to threshold selection." *Extremes*, 1(3): 251–261.

Efron, B. 1981. "Nonparametric standard errors and confidence intervals." *Canadian Journal of Statistics*, 9(2): 139–158.

Elliott, Robert J, and Hong Miao. 2009. "VaR and expected shortfall: a non-normal regime switching framework." *Quantitative Finance*, 9(6): 747–755.

Estep, Tony, and Mark Kritzman. 1988. "TIPP: Insurance without complexity." *The Journal of Portfolio Management*, 14(4): 38–42.

Fama, E.F. 1984. "Forward and spot exchange rates." *Journal of Monetary Economics*, 14(3): 319–338.

Fama, E.F., and K.R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33(1): 3–56.

Fama, E.F., and R.R. Bliss. 1987. "The information in long-maturity forward rates." *The American Economic Review*, 77(4): 680–692.

Fuertes, A.M., J. Miffre, and G. Rallis. 2010. "Tactical allocation in commodity futures markets: Combining momentum and term structure signals." *Journal of Banking & Finance*, 34(10): 2530–2548.

Fung, W., and D.A. Hsieh. 2000. "Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases." *Journal of Financial and Quantitative Analysis*, 35(3): 291–308.

Fung, W., and D.A. Hsieh. 2001. "The risk in hedge fund strategies: Theory and evidence from trend followers." *Review of Financial Studies*, 14(2): 313–341.

Fung, W., and D.A. Hsieh. 2004. "Hedge fund benchmarks: A risk-based approach." *Financial Analysts Journal*, 60(5): 65–80.

Fung, W., D.A. Hsieh, N.Y. Naik, and T. Ramadorai. 2008. "Hedge funds: Performance, risk, and capital formation." *The Journal of Finance*, 63(4): 1777–1803.

Furnival, G.M., and R.W. Wilson Jr. 1974. "Regressions by leaps and bounds." *Technometrics*, 16(4): 499–511.

Getmansky, M., A.W. Lo, and I. Makarov. 2004. "An econometric model of serial correlation and illiquidity in hedge fund returns." *Journal of Financial Economics*, 74(3): 529–609.

Giot, Pierre, and Sébastien Laurent. 2004. "Modelling daily value-at-risk using realized volatility and ARCH type models." *journal of Empirical Finance*, 11(3): 379–398.

Gisclon, P. 2011. "Dynamic Trading Strategies of Hedge Funds: Implications for Risk and Return." PhD diss. No. 3894, University of St.Gallen.

Gorton, G.B., F. Hayashi, and K.G. Rouwenhorst. 2007. "The fundamentals of commodity futures returns." National Bureau of Economic Research.

Haas, Markus, Stefan Mittnik, and Marc S Paolella. 2004. "Mixed normal conditional heteroskedasticity." *Journal of Financial Econometrics*, 2(2): 211–250.

Hampel, FR, EM Ronchetti, and PJ Rousseeuw. 1986. "Robust statistics: the approach based on influence functions." *Wiley series in probability and mathematical statistics*.

Hansen, Bruce E. 1994. "Autoregressive conditional density estimation." *International Economic Review*, 35(3): 705–730.

Hansen, Peter R, and Asger Lunde. 2005. "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" *Journal of applied econometrics*, 20(7): 873–889.

Hastie, T., R. Tibshirani, and J. Friedman. 2004. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." *BeiJing: Publishing House of Electronics Industry*.

Hou, K., and M. Van Dijk. 2010. "Profitability shocks and the size effect in the cross-section of expected stock returns." *Working Paper*.

Huber, O. 2011. "Investigating Hedge Fund Performance." PhD diss. University of St.Gallen.

Jegadeesh, N., and S. Titman. 1993. "Returns to buying winners and selling losers: Implications for stock market efficiency." *Journal of finance*, 48(1): 65–91.

Jensen, M.C. 1968. "The performance of mutual funds in the period 1945-1964." *The Journal of Finance*, 23(2): 389–416.

Kroencke, T., F. Schindler, and A. Schrimpf. 2011. "International Diversification Benefits with Foreign Exchange Investment Styles." *Working Paper. Center for European Economic Research (ZEW).*

Kuester, Keith, Stefan Mittnik, and Marc S Paolella. 2006. "Value-at-risk prediction: A comparison of alternative strategies." *Journal of Financial Econometrics*, 4(1): 53–89.

Kupiec, Paul. 1995. "Techniques for verifying the accuracy of risk measurement models." *The Journal of Derivatives*, 3(2): 73–84.

Lattimore, Pamela K, Joanna R Baker, and Ann D Witte. 1992. "The influence of probability on risky choice: A parametric examination." *Journal of economic behavior & organization*, 17(3): 377–400.

Leeb, H., and B.M. Pötscher. 2005. "Model selection and inference: Facts and fiction." *Econometric Theory*, 21(1): 21–59.

Liang, B. 1999. "On the performance of hedge funds." *Financial Analysts Journal*, 55(4): 72–85.

Lustig, H., N. Roussanov, and A. Verdelhan. 2011. "Common Risk Factors in Currency Markets." *Review of Financial Studies*, 24(11): 3731–3777.

Mancini, Loriano, and Fabio Trojani. 2011. "Robust Value at Risk Prediction." *Journal of Financial Econometrics*, 9(2): 281–313.

Mancini, Loriano, Elvezio Ronchetti, and Fabio Trojani. 2005. "Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models." *Journal of the American Statistical Association*, 100(470): 628–641.

McNeil, Alexander J, and Rüdiger Frey. 2000. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach." *Journal of empirical finance*, 7(3): 271–300.

Meligkotsidou, L., and I.D. Vrontos. 2008. "Detecting structural breaks and identifying risk factors in hedge fund returns: A Bayesian approach." *Journal of Banking & Finance*, 32(11): 2471–2481.

Miffre, J., and G. Rallis. 2007. "Momentum strategies in commodity futures markets." *Journal of Banking & Finance*, 31(6): 1863–1886.

Mittnik, Stefan, and Marc S Paolella. 2000. "Conditional density and value-at-risk prediction of Asian currency exchange rates." *Journal of Forecasting*, 19(4): 313–333.

Naik, N.Y., T. Ramadorai, and M. Stromqvist. 2007. "Capacity constraints and hedge fund strategy returns." *European Financial Management*, 13(2): 239–256.

Newey, Whitney K, and Kenneth D West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): 703–08.

Newey, Whitney K, and Kenneth D West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, 61(4): 631–53.

Pojarliev, M., and R.M. Levich. 2008. "Do Professional Currency Managers Beat the Benchmark?" *Financial Analysts Journal*, 64(5): 18–32.

Pötscher, B.M. 1991. "Effects of model selection on inference." *Econometric Theory*, 7(2): 163–185.

Pritsker, Matthew. 2006. "The hidden dangers of historical simulation." *Journal of banking & finance*, 30(2): 561–582.

Queisser, Monika, and Dimitri Vittas. 2000. "The Swiss multi-pillar pension system: triumph of common sense?" *World Bank Policy Research Working Paper*, 2416.

Rubinstein, Mark, and Hayne E. Leland. 1981. "Replicating options with positions in stock and cash." *Financial Analysts Journal*, 37(4): 63–72.

Sandvik, S.H., S. Frydenberg, S. Westgaard, and R.K. Heitmann. 2011. "Hedge Fund Performance in Bull and Bear Markets: Alpha Creation and Risk Exposure." *The Journal of Investing*, 20(1): 52–77.

Sbaraglia, S, M Papi, M Briani, M Bernaschi, and F Gozzi. 2003. "A model for the optimal asset-liability management for insurance companies." *International Journal of Theoretical and Applied Finance*, 6(03): 277–299.

Schulze, Isabelle, et al. 2007. *The handbook of West European pension politics.* Oxford University Press.

Shiller, Robert J. 1992. *Market volatility.* The MIT Press.

Taylor, James W. 2008*a*. "Estimating value at risk and expected shortfall using expectiles." *Journal of Financial Econometrics*, 6(2): 231–252.

Taylor, James W. 2008*b*. "Using exponentially weighted quantile regression to estimate value at risk and expected shortfall." *Journal of Financial Econometrics*, 6(3): 382–406.

Titman, S., and C. Tiu. 2011. "Do the best hedge funds hedge?" *Review of Financial Studies*, 24(1): 123–168.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty*, 5(4): 297–323.

Victoria-Feser, Maria-Pia, and Elvezio Ronchetti. 1994. "Robust methods for personal-income distribution models." *Canadian Journal of Statistics*, 22(2): 247–258.

Vigna, Elena, and Steven Haberman. 2001. "Optimal investment strategy for defined contribution pension schemes." *Insurance: Mathematics and Economics*, 28(2): 233–262.

Wilcox, Rand R. 2012. *Introduction to robust estimation and hypothesis testing.* Academic Press.

Yamai, Y., and T. Yoshiba. 2001. "Comparative analyses of expected shortfall and value-at-risk (2): expected utility maximization and tail risk." *Monetary and economic studies*, 20(2): 95–115.

Zhong, Z. 2008. *Why does hedge fund alpha decrease over time? Evidence from individual hedge funds.* ProQuest.

Zhu, D., and J.W. Galbraith. 2011. "Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions." *Journal of Empirical Finance*, 18(4): 765–778.

Zhu, Dongming, and John W Galbraith. 2010. "A generalized asymmetric Student-t distribution with application to financial econometrics." *Journal of Econometrics*, 157(2): 297–305.

Zieling, Daniel, Antje Mahayni, and Sven Balder. 2014. "Performance evaluation of optimized portfolio insurance strategies." *Journal of Banking & Finance*, 43: 212–225.

Curriculum Vitae

Thomas Rolf Althaus, born on August 1, 1983, in Zürich, Switzerland

Education

2011-2014	Ph.D. in Management, University of St.Gallen
2006-2008	Master of Arts in Quantitative Economics and Finance, University of St.Gallen
2007	Exchange Semester, University of Texas at Austin
2003-2006	Bachelor of Arts in Business Administration, University of St.Gallen
1998-2002	Matura, Kantonsschule am Burggraben, St.Gallen

Work Experience

2014	Wiederkehr Associates AG, Chief Investment Officer/Managing Partner
2011-2013	Arecon AG, Portfolio Manager/Head Risk Management
2010	Goldman Sachs International, Analyst
2008-2010	LGT Capital Management Ltd., Quantitative Portfolio Manager
2003-2005	Charles Vögele AG, Sales Assistant
2002-2003	Mercer Management Consulting, Assistant