Year: 2017

Three essays in portfolio optimization

Adelmann, Maximilian

Three Essays in Portfolio Optimization

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Maximilian Adelmann
from Germany

approved in February 2017 at the request of

Prof. Dr. Karl Schmedders
Prof. Dr. János Mayer

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, February 15, 2017

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

Acknowledgements

I am heavily indebted and very grateful to my supervisors János Mayer and Karl Schmedders for their teaching, guidance, and support on this project. Also, I would like to express my special thanks to Lucio Arjona Fernandez and Regine Scheder for helpful insights into replicating portfolios in practice and discussions on the subject. Finally, I thank Lorenz Beyeler and my co-author Cyril Bachelard for the great collaboration.

Contents

Part I

Introduction

General Introduction

The importance of portfolio optimization models in practice has increased a great deal with advances in computational power. Nowadays they are used for various areas in the financial industry. The general idea behind portfolio optimization models is the determination of assets holdings that maximize or minimize a certain criterion. The most well-known of these models is certainly the mean-variance model developed by Markowitz (1952). In this intuitive approach an investor determines the assets holdings that maximize (minimize) the expected portfolio mean (variance) in the presence of certain constraints. Numerous other portfolio optimization models build on the work of Markowitz (1952); all serve the main purpose of maximizing the portfolio performance of an investor and are, for example, used by banks and asset management companies. However, investment maximization is not the only application of portfolio optimization models in practice. They can also serve as an important tool for the valuation of companies' liabilities, for example in the insurance industry.

In this thesis we analyze portfolio optimization models with a practical relevance. For each chapter we implement our models using several real-life, large-scale data sets. In the first two chapters we examine optimization models that aim to maximize an investor's performance. In Chapter 1 we discuss the choice of a successful portfolio optimization model and develop a model that combines the strengths of the global minimum variance (GMV) portfolio by Markowitz (1952) and another well-known model, developed by Black and Litterman (1990). Chapter 2 addresses one of the main issues in the portfolio optimization field: the estimation of models' input parameters. We develop a forecasting approach for predicting probabilities for returns larger than certain percentiles. Based on these forecasts we conduct a trading strategy in which an investor has the choice to invest in the risk-free rate or the stock market. In Chapter 3 we examine an optimization problem in a different portfolio optimization field: the valuation of companies' liabilities. In particular we focus on the life insurance industry and develop a replicating portfolio (RP) model that approximates life insurance liabilities as closely as possible.

We address the topics in the following way:

Chapter 1 addresses the choice of a portfolio optimization model for asset management companies. An asset management company faces two potentially contradictory goals. On the one hand, a portfolio optimization model has to be intuitive and easily understandable if it is to be sold to potential clients. On the other hand, the model should have a small risk exposure and provide high returns on investments. In practice, many asset management companies nowadays still use optimization models based on the mean variance approach developed by Markowitz (1952), despite its well-known shortcomings, including, for example, the difficulty to estimate expected returns for all assets and the high sensitivity to input parameters. One portfolio in the Markowitz framework that overcomes these weaknesses of the traditional mean variance approach is the GMV portfolio. For this portfolio we do not need mean estimations, which are usually unreliable. Instead, the covariance matrix—which is simpler to estimate by comparison—is a sufficient input. Another portfolio optimization model that has generated high returns in practice is the Black–Litterman (B–L) model. This model—introduced by Black and Litterman (1990)—combines the market equilibrium with investors' views on the performance of some (but usually not all) assets in the universe. This model was implemented successfully in practice at Goldman Sachs. However, its shortcomings are its high complexity and a high degree of uncertainty with regard to its input parameters.

In Chapter 1 we develop a model that combines the GMV portfolio with the B–L model to reap the benefits of both approaches. In contrast to the original B–L approach we use the intuitive GMV portfolio as the reference portfolio. Furthermore, we introduce a general rule to simulate investors' views to improve the GMV portfolio by clearing it of so-called dead assets. Our approach mostly relies on covariance estimations. Mean estimations only have a minor impact for the simulation of investors' views in our general rule. Thereby we only consider relative values of expected returns between assets. A numerical application of our model to empirical data sets demonstrates that portfolios based on the model clearly outperform the GMV portfolio and the $1/N$ portfolio in terms of compound annual returns and Sharpe ratios.

Chapter 2 focuses on an essential portfolio optimization issue: the forecast of input parameters. In Chapter 1 we point out that mean estimations are unreliable. However, Christoffersen and Diebold (2006) show that the sign of asset returns can be forecasted despite the absence of mean "forecastability". The idea behind sign prediction is to provide a probability—expressed as a percentage—for a return larger than zero in the next period. Sign predictions are based on explanatory variables, which can—in contrast to the mean—be forecasted. These explanatory variables include, for example, volatility (see Christoffersen and Diebold (2006)), skewness and kurtosis (see Christoffersen et al. (2007)), the interest rate and the term spread (see Chevapatrakul (2013)), or a recession indicator (see Nyberg (2011)). A suitable choice for sign prediction is classification models, in particular logit and probit approaches. Methods for evaluating predictive power differ in the literature. Common forecasting tests include market timing tests (see Pesaran and Timmermann (1992) and Anatolyev and Gerko (2005)), Brier scores (see Christoffersen et al. (2007), Anatolyev (2009), and Nyberg and Pönkä (2015)), and pseudo-R^2 measures (see Chevapatrakul (2013) and Nyberg and Pönkä (2015)). Sign predictions can be applied in trading strategies. In these strategies, investments are made in the stock market or a safer alternative—such as bonds or treasury bills—based on the probability of a positive return from the stock market in the next period.

In Chapter 2 we take the idea and intuition of sign forecasting one step further and extend it to also forecast the probabilities of returns larger than certain percentiles. Such percentile forecasts can be particularly useful for an investor who wants to avoid periods with large negative returns and to reduce the volatility of his or her investments. Our prediction model is based on a binary logit approach and we forecast probabilities for returns larger than 0%, −2%, −4%, −6%, and −8%. As explanatory variables we use the conditional volatility calculated with a GARCH(1,1) approach, the unconditional skewness, and the Mahalanobis distance (a turbulence measure that has not been used in the sign forecasting literature before). We evaluate our percentile forecasts with absolute Brier scores, a pseudo-R^2 measure, and a new evaluation test, which is well suited to a visualization of predictive power. Test results show a superior forecasting performance of models that include the Mahalanobis distance and forecasted percentiles smaller than 0%. In the last step of our analysis we conduct a trading strategy based on our percentile forecasts. Depending on the probabilities of all asset forecasts we invest either in the stock market or the risk-free rate. Investments based on such a trading strategy provide better Sharpe ratios than their respective benchmark strategies.

Chapter 3 addresses a different field of portfolio optimization than the two previous chapters. Instead of models with an investment purpose this chapter focuses on portfolio approaches for the valuation of companies' liabilities. In particular we analyze the application and benefits of an RP model for the life insurance industry, which faces dramatic changes due to the Solvency II directive. This directive—a new regulatory framework for the European Union—has established new capital requirements, valuation techniques, and governing and reporting standards for insurance companies. It includes quantitative requirements, which force insurance companies to calculate the "fair" value of their liabilities. Some of these liabilities—such as, for example, the values of life insurance policies—are long-term and dependent on market conditions. An accurate valuation of such liabilities is very complex and—thus—several different evaluation techniques have become popular recently, among them approaches based on RPs. An RP (in the insurance context) is a pool of a finite number of selected financial instruments designed to (approximately) reproduce the cash flows or present values of liabilities across a large number of economic scenarios. Its optimization process comprises three steps. First, the determination of future scenario sets and the simulation of liabilities and candidate assets for different scenarios within these scenario sets. Second, the optimization of a portfolio—based on the candidate assets—that replicates the liabilities as closely as possible based on a certain metric. Third, a quality assessment of the RP in the light of several tests.

In Chapter 3 we focus on the second and the third steps of the RP optimization process. We take a set of candidate assets and cash flow vectors for liabilities and assets as given. We then formulate an L_1 norm (which is superior to the L_2 norm for the detection of outliers) optimization model to minimize the absolute difference between the liability cash flows and our RP cash flows. We employ two different linear reformulations of the L_1 minimization problem. This enables us to solve our RP optimization model over many years and to include computationally cheap additional constraints. Our sets of additional constraints are supposed to ensure a relatively small RP and an RP that will likely perform in the same way as the mapped liabilities in a large range of different future scenarios. We assess the performance of our RP with several tests that are all performed with out-of-sample data. In particular we focus on the out-of-sample movements of the scenario sets and the market value fit between liability cash flows and those of the RP. Furthermore, we calculate the coefficient of determination, R^2, between the RP and the liability cash flows. A numerical analysis demonstrates that our model delivers RPs with excellent practical properties in a reasonable amount of time.

Part II

Three Essays in Portfolio Optimization

Essay 1

Minimum Variance and Black–Litterman

An Improvement of the Global Minimum Variance Portfolio using a Black–Litterman Approach[1]

Maximilian Adelmann
Dept. of Business Administration
University of Zurich
Moussonstrasse 15
8044 Zurich, Switzerland
maximilian.adelmann@business.uzh.ch

May 2016

Abstract

Asset management companies are constantly searching for portfolio optimization models that are on the one hand clear and intuitive and on the other provide high and reliable returns. This paper presents a modified version of the well-known Black–Litterman portfolio optimization approach. Unlike in the original model, the intuitive global minimum variance portfolio (GMV) serves as the reference portfolio. The introduction of a general rule for investors' views in combination with a simplification of the original Black–Litterman approach facilitates the implementation of the model and enables us to remove so-called dead assets from the GMV portfolio. As an additional advantageous feature our model is only based on variance-covariance estimations, and relative return estimations for our general rule. A numerical application of our modified Black–Litterman model to empirical data sets demonstrates that portfolios based on the model clearly outperform the GMV portfolio and the $1/N$ portfolio in terms of compound annual returns and Sharpe ratios.

Keywords: Portfolio optimization, Black–Litterman model, global minimum variance portfolio, back-testing.

[1] I thank Lorenz Beyeler, Cyril Bachelard, and Carmine Orlacchio for helpful discussions on the subject and comments on earlier implementations of the model. I am grateful to seminar audiences at OLZ & Partners, the Hoover Institution at Stanford, and the World Finance Conference 2016 in New York for helpful comments. I would also like to thank my supervisors János Mayer and Karl Schmedders for their support and guidance on this project.

1.1 Introduction

Asset management companies are constantly searching for portfolio optimization models that are on the one hand clear and intuitive and on the other provide high and reliable returns. This paper presents a modified version of the well-known Black–Litterman (B–L) portfolio optimization approach. Unlike in the original model, the intuitive global minimum variance (GMV) portfolio serves as the reference portfolio. The introduction of a general rule for investors' views in combination with a simplification of the original B–L approach facilitates the implementation of the model and enables us to remove so-called dead assets from the GMV portfolio. As an additional advantageous feature our model is only based on variance-covariance estimations, and relative return estimations for our general rule. A numerical application of our modified B–L model to empirical data sets demonstrates that portfolios based on the model clearly outperform the GMV portfolio and the $1/N$ portfolio in terms of compound annual returns and Sharpe ratios.

The quest for a successful portfolio optimization model is one of the main challenges for an asset management company. Such a company has to find the right balance between two potentially contradictory goals. On the one hand a portfolio optimization model has to be clear, understandable, and intuitive if it is to be sold to potential clients. On the other hand the model has to be reliable and low-risk and provide a high return on investments.

One portfolio optimization model that is still widely used at asset management companies today is the mean-variance model developed by Markowitz (1952). Relying purely on the trade-off between expected return and variance, the model is highly intuitive, which makes it attractive for potential clients. However, it also has several shortcomings, including the difficulty of estimating expected returns for all assets and a high sensitivity to input parameters. The mean-variance model therefore often does not perform well in practice and is sometimes even outperformed by simple allocation strategies such as the $1/N$ portfolio strategy (see De Miguel et al. (2009)).

Asset management companies—such as OLZ & Partners Asset and Liability Management AG (OLZ) in Bern, Switzerland—try to overcome the weakness of the standard mean-variance approach by using the GMV optimization model to determine the holdings of their portfolio. The advantage of this approach is that no estimators of expected returns are needed. This is an advantageous feature because the estimation of expected returns is complicated and results are usually not very reliable. The estimation of the variance-covariance matrix—which is the only required input for the GMV optimization—is simpler in comparison to the estimation of expected returns and the results obtained are much more reliable. However, the disadvantage of this approach is the complete absence of return estimations in the portfolio optimization process. According to OLZ this could result in unfavorable portfolios including several so-called dead assets. These dead assets have (small) negative returns in almost every period. Metaphorically speaking, they slowly "die". These small negative returns in almost every period have the effect of a negative correlation of all these assets with the "normal" assets in the portfolio, which usually have positive returns most of the time. In the setup of the GMV approach these small correlation values lead to portfolios with significant holdings in dead assets. This leads to the returns of these portfolios being reduced by the inclusion of these dead assets.

One portfolio optimization model that has generated high returns in practice is the B–L model. The basic idea of the model—the combination of the market equilibrium with investors' views—was introduced by Black and Litterman (1990). The model was successfully implemented at Goldman Sachs. However, its shortcomings include its high complexity—compared to, for example, the mean-variance model—and the high degree of uncertainty with regards to its input parameters. Several papers have tried to explain the model in detail or to demystify it, and these include Satchell and Scowcroft (2000) and Idzorek (2005). But certain of its aspects, such as parameter values, are still not completely clear.

In our portfolio optimization model we use the GMV approach applied by OLZ as a starting point. To avoid dead assets in the portfolio we combine the GMV approach with a modified B–L approach. In our model the GMV portfolio—rather than the market equilibrium portfolio—serves as the reference portfolio in the B–L setup. Then we use a general rule to simulate investors' views in the B–L setup to improve the GMV portfolio by clearing it of dead assets. Our approach mostly relies on variance-covariance estimations such as the simple GMV approach. We only use return estimates for the general rule in our B–L model, which simulates investors' views. Thereby we only consider relative values of expected returns between assets. This drastically reduces the negative impact of errors in return estimations compared to the absolute values that are used in the classic mean-variance approach. In general our model combines the benefits of both the GMV and the B–L model. On the one hand we have the intuitive and clear GMV model, which is rather easy to implement. On the other we enrich the GMV model with valuable features of the B–L setup. Thereby our general rule overcomes implementation problems such as certain of the parameter specifications usually associated with the original B–L model.

In this paper we first briefly explain the original B–L approach. We show how the market equilibrium, as a reference portfolio, gets adjusted by investors' views. Next we present our new B–L approach to improving the GMV portfolio. Therefore, we first recall the calculation of the GMV portfolio with a non-negativity constraint. Then we present a modified B–L approach with this GMV portfolio as the reference portfolio. In our modified approach we also use a non-negativity constraint—that is to say, short sales are excluded.

In a second step we introduce a guideline for the practical implementation of our model. We describe the two real-life, large-scale data sets and the back-testing approach that we used for the implementation. We then explain a general rule for simulating investors' views in detail, which is a novelty with regards to the existing B–L literature. With this general rule we create views for certain assets based on the systematic risk β and the expected return of all assets. As a result of this general rule we facilitate our B–L approach by assuming 100 percent confidence in our views. This enables us to remove two critical parameters from our equations.

We present detailed results of our implementation for several portfolios created with our model. We asses the quality of our model using the following out-of-sample performance criteria: the cumulative return (CR), the compound annual return (CAR), and the Sharpe ratio. Thereby the GMV portfolio and the $1/N$ portfolio serve as benchmarks. We also provide a thorough sensitivity analysis for other parameters of our model.

The Literature

Modern portfolio theory started with Markowitz (1952) paper on portfolio selection. The strength of this approach is a brilliant quantification of expected return maximization and risk minimization as two basic investment objectives (see He and Litterman (1999)). However, the traditional mean-variance approach has several well-known shortcomings. Mean-variance portfolios often do not perform well in practice and are in some cases even outperformed by simple allocation strategies such as the $1/N$ portfolio strategy (see De Miguel et al. (2009)). Furthermore mean-variance portfolios have usually extreme positions and are not particularly intuitive (see He and Litterman (1999)). One main reason for the undesirable results of mean-variance portfolio use in practice is the difficulty of estimating expected returns (see Black and Litterman (1992)). The Markowitz formulation unrealistically requires specific values for each title in the asset universe (He and Litterman (1999)). The other main reason for undesirable results in practice is the very high sensitivity of the mean-variance model to small changes in the input, especially the return estimators (see Fabozzi (2008) and He and Litterman (1999)). Best and Grauer (1991) for example show that a marginal change in the mean of a single asset can result in a change of half of the portfolio composition. This leads basically to an optimization of errors and very unstable portfolios (see Mankert (2010)).

A special case in the Markowitz framework is the GMV portfolio. This portfolio, with the smallest possible volatility, does not consider estimated returns and only needs variance-covariance estimations as inputs. Since the variance-covariance estimation is more accurate than the mean estimation from a series of realized returns (Merton (1980)) and the impact of errors in means is much larger on the optimal portfolio than the impact of errors in variances and covariances, this is a desirable feature for a portfolio optimization process (see Chopra and Ziemba (1993)). Despite its rather simple input requirements the GMV portfolio often outperforms portfolios based on more advanced optimization models (see Jorion (2000)). However, the GMV portfolio also has a few shortcomings. First, GMV portfolios are usually under-diversified (see De Miguel et al. (2009)). Second, ignoring expected returns completely is a waste of potentially important information.

A different portfolio optimization approach—also developed to overcome weaknesses in the Markowitz framework—is the B–L model. The basic idea of this model—combining the market equilibrium with investors' views—was introduced by Black and Litterman (1990). Black and Litterman (1992) expanded their framework to the subject of global portfolio optimization to provide a quantitative asset allocation model that is relatively easy to use and well behaved. He and Litterman (1999) illustrate the intuition behind the B–L asset allocation models, using examples to illustrate differences between the Markowitz and the B–L frameworks.

Several authors explain the B–L model in detail or present implementation examples to facilitate an understanding of the model. Idzorek (2005) provides a step-by-step guide using a simple example with eight assets. Mankert (2010) derives the B–L approach in great detail and fills some knowledge gaps concerning certain parameters. Mankert and Seiler (2011) provide a more compact version of the mathematical derivations of the B–L model and practical implications for its use. Cheung (2009) explains the model focusing on an economic interpretation, a clarification of assumptions, and an implementation guidance. Walters (2014) provides a complete description of the B–L model including full derivations from its underlying principles. He also considers the various parameters of the model and existing key papers in the B–L framework. Satchell and Scowcroft (2000) present several examples of B–L asset allocation models and highlight the combination of judgmental and quantitative views.

A few papers provide interesting extensions to the B–L model. Idzorek (2005) introduces a new method to include the investor's confidence in his views to the model. Fabozzi (2008) incorporate factor models and cross-sectional rankings into the B–L framework. Meucci (2010) provides minor modifications of the original B–L model to improve the range of its applications. Bertsimas et al. (2012) replace the statistical B–L framework with ideas from inverse optimization, expanding the scope and applicability of the model.

The B–L model was originally implemented at Goldman Sachs in the 1990s. Bevan and Winkelmann (1998) provide a summary of their practical experience of using the model for their investment strategy at the firm. They show how they set the key parameters and determined their views and the confidence in those views. On a much smaller scale and with certain adjustments the B–L model was also implemented at a Swedish Bank (see Mankert (2010)).

The original B–L model is a combination of the market equilibrium portfolio, which serves as a reference portfolio, and certain user-specific views. However, according to Haesen et al. (2014) the reference portfolio does not necessarily have to be the market equilibrium portfolio. They present a variation of the model with the risk-parity portfolio as a starting point.

In general this paper makes three main contributions to the existing state of research into BL. First we enrich the theory by creating a modified model with the GMV portfolio as the reference portfolio. This combination of two popular models enables us to eliminate one of the biggest weaknesses of the GMV portfolio (the total absence of mean estimates) and avoid the main drawback of the classic mean-variance model (the strong negative impact of errors in mean estimation). Second we introduce a general rule to test the views of our modified B–L model in a back-testing context. This general rule facilitates the implementation of the model in practice and enables one to assess the model's quality out-of-sample. Third we analyze the practical suitability of our model with two large-scale, real-life data sets. The size of these data sets (they contain 591 and 940 assets, respectively) ensures a more realistic analysis than existing studies based on small data sets that contain usually between 10 and 50 assets. This leads to new, valuable theoretical and practical insights.

The remainder of this paper is organized as follows. We introduce our modified B–L model in Section 1.2. Section 1.3 presents the implementation of our model, including a description of the data sets, our back-testing approach, a general rule to simulate investors' views, and the specific problem instance. Section 1.4 contains the main results of the paper. In particular, Section 1.4.1 compares the out-of-sample returns of different portfolios and Section 1.4.2 provides compound annual returns, standard deviations, and Sharpe ratios for these portfolios. In Section 1.4.3 we show the effect of variations of certain parameters of our model. Section 1.4.4 provides the composition of our modified B–L portfolios. Section 1.5 concludes.

1.2 The Model

In a first step we describe and explain the original B–L approach as it serves as the basis for our own model. In a second step we show the modifications that we make to use the B–L framework to improve the GMV portfolio.

1.2.1 The Original Black–Litterman Approach

Table 1.1: Black–Litterman Parameters.

Symbol	Description	Size
n	Number of available assets	-
k	Number of views	-
Π	Implied excess equilibrium return vector	$n \times 1$
δ	Risk aversion coefficient	1×1
Σ	Estimated covariance matrix	$n \times n$
w_M	CAPM market weights	$n \times 1$
μ^*	New combined return vector	$n \times 1$
P	Asset weights involved in the views	$k \times n$
q	View vector	$k \times 1$
τ	Weight-on-views	1×1
Ω	Uncertainty in each view	$k \times k$
w	B–L model portfolio weights	$n \times 1$

This table provides a parameter overview for the original B–L approach (see Idzorek (2005)).

The B–L model was originally designed to enable investors to include their views about certain assets in the market equilibrium model in order to achieve intuitive, diversified portfolios.[2] The starting point of the B–L model is equilibrium returns, which are often taken from the capital asset pricing model (CAPM). As does Idzorek (2005) we first calculate the implied excess equilibrium return vector Π:

$$\Pi = \delta \Sigma w_M, \tag{1.1}$$

where δ (sometimes also denoted as λ in the literature) is the risk aversion factor, Σ is the covariance

[2] In this subsection we follow the representation of the B–L approach in Idzorek (2005) and Mankert (2010). For the original B–L global portfolio optimization setup see Black and Litterman (1992).

matrix of the excess returns, and w_M are the market weights from the CAPM.

In the next step we introduce some matrices and vectors to express the views of an investor (see Idzorek (2005)). We have k views with $k \leq n$, because we can have at most one view per asset. We use a matrix P to express which assets are involved in the views. This matrix has one row for each view k and one column for each available asset n. We have a vector q that we use to express the investors' views. This is a column vector consisting of k elements. In these elements we express our views about the expected return. We use absolute views for these expected returns and not relative views compared to the implied excess equilibrium return vector. Our matrix Ω is a diagonal covariance matrix of error terms from the expressed views representing the uncertainty in each view (see Idzorek (2005)). For the special case of 100 percent certainty in each view this matrix would only consist of zeros. For all the other cases it has k rows and k columns. Finally, we have a scalar τ, which determines the weight-on-views compared to the weights based on the CAPM. Table 1.1 provides an overview of all parameters.

These parameters enable us to calculate the new combined return vector μ^*, consisting of n rows, with the following equation (see Idzorek (2005)):

$$\mu^* = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}q]. \tag{1.2}$$

This equation combines the results of the equilibrium model—which is usually the CAPM—with the investor's views. A disadvantage of this formulation is the need for several computationally expensive matrix inversions of a large $n \times n$ matrix. Mankert (2010) shows how this equation may be reformulated:[3]

$$\mu^* = \Pi + \tau\Sigma P'(\Omega + \tau P\Sigma P')^{-1}(q - P\Pi). \tag{1.3}$$

The advantage of the new formulation is that only a $k \times k$ matrix inversion is required, which is computationally relatively cheap.

The literature—see for example Mankert (2010) and Idzorek (2005)—provides two ways of determining the optimum weights w based on the new combined return vector μ^*. In the first way, an optimization of the new combined return vector μ^*, carried out in a mean-variance manner, is used to determine the optimal weights w of the portfolio

$$\max_w w'\mu^* - (\delta w'\Sigma w)/2. \tag{1.4}$$

As a second way of finding w, Idzorek (2005) and Mankert (2010) suggest rearranging (1.1) and substituting μ^* for Π

$$w = (\delta\Sigma)^{-1}\mu^*. \tag{1.5}$$

The results for w are the same for both ways.

1.2.2 A Black–Litterman Approach to Improving the GMV Portfolio

In the next step we present our own model, which combines the mean-variance optimization model with the B–L approach. First we describe the GMV optimization model, which serves as the reference

[3] For a proof that (1.2) and (1.3) are equivalent see Mankert (2010).

portfolio for a slightly modified B–L approach, which we present second.

1.2.2.1 GMV Portfolio Optimization

Table 1.2: Parameters of the Model.

Symbol	Description	Size
n	Number of available assets	-
k	Number of views	-
Π_N	Implied excess reference return vector	$n \times 1$
δ	Risk aversion coefficient	1×1
Σ	Estimated covariance matrix	$n \times n$
w_N	GMV portfolio weights	$n \times 1$
μ^*	New combined expected return vector	$n \times 1$
P	Assets involved in the views	$k \times n$
q	View vector	$k \times 1$
τ	Weight-on-views	1×1
Ω	Uncertainty in each view	$k \times k$
w	B–L model portfolio holdings	$n \times 1$

This table provides a parameter overview for our modified B–L approach.

In a first step we minimize the variance of the portfolio according to Markowitz (1952) using n potential assets

$$\min_{w} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_i \Sigma_{ij} w_j), \tag{1.6}$$

where the decision variables w_i are the weights of the holdings in each of the i^{th} potential assets and Σ is the variance-covariance matrix of the potential assets. We face the constraint that the weights w_j have to sum up to 1

$$\sum_{j=1}^{n} w_j = 1. \tag{1.7}$$

Many asset management companies—including OLZ—limit themselves to portfolios without short sales. Therefore we focus on a Markowitz framework without shorting and introduce the following non-negativity constraint:

$$w_j \geq 0 \quad \forall\ j. \tag{1.8}$$

1.2.2.2 Modification of the Black–Litterman Approach

For our model we slightly modify the B–L approach. Table 1.2 provides a parameter overview for this modified approach. The first and most essential modification takes place at the starting point of the model—the calculation of the implied excess equilibrium return vector Π. Litterman (2003) recommends the equilibrium approach as an appropriate reference point. He does not define it as an unchangeable requirement though. Based on the discussions in Litterman and the Quantitative Resources Group (2003), Mankert (2010) illustrates the equilibrium concept of the B–L model, stating that "the idea of an equilibrium as a point of reference for the B–L model is a kind of ideal condition for the model." However, for an application to real life investments, reasonable approximations of this ideal state have, according to Mankert (2010), to be made. Haesen et al. (2014) are also of the opinion that the reference portfolio does not necessarily have to be equal to the market equilibrium portfolio. For our model we replace the implied excess equilibrium return vector Π by the implied excess reference return vector Π_N. This vector is calculated with the weights w_N of the GMV portfolio instead of the market weights w_M:

$$\Pi_N = \delta \Sigma w_N. \tag{1.9}$$

Using the GMV portfolio instead of an equilibrium portfolio—such as the CAPM portfolio—is a deviation from the ideal condition of the model. However, in a real life setting with large-scale portfolios it is still a reasonable approximation of a reference point to determine the implied views.

In the next step we calculate our new combined return vector μ^*. The formula is almost the same as in Equation (1.3). The only difference is that we use Π_N instead of Π:

$$\mu^* = \Pi_N + \tau \Sigma P'(\Omega + \tau P \Sigma P')^{-1}(q - P\Pi_N). \tag{1.10}$$

In the last step we use the optimization approach shown in (1.4) to determine our new optimal portfolio holdings w^*:

$$\max_w w'\mu^* - (\delta w' \Sigma w)/2. \tag{1.11}$$

To avoid negative values for our portfolio holdings we have to include a non-negativity constraint on our optimization problem:

$$w_j \geq 0 \quad \forall\ j. \tag{1.12}$$

In a mean-variance setup a portfolio optimization problem for determining the holdings of the model—such as (1.11)—would contain a normalization constraint. Our numerical experiments show that such a constraint leads in the case of our model to poorly diversified portfolios with a bad out-of-sample performance. However, having solved (1.11) we need the sum of the optimal holdings w_i^* to be equal to 1 such as in the Markowitz framework to ensure comparability to the GMV portfolio. Therefore we normalize the optimal holdings after the optimization by dividing w_i^* by the sum of all holdings w_j^* (under the assumption that the sum of all holdings w_j^* is not zero) to get our portfolio weights w_{if}^*:

$$w_{if}^* = w_i^* / \sum_{j=1}^{n} w_j^* \quad \forall\ i. \tag{1.13}$$

1.3 Implementation of the Model

We implement our model according to the formulas developed in Section 1.2.2.1 and Section 1.2.2.2. We use two different data sets for the implementation, the MSCI World index and the S&P EMBI.[4] To assess the out-of-sample performance of our model we use a back-testing approach. We determine the input parameters for our specific problem instance according to the B–L literature. The creation of certain specific views is complicated in a back-testing context. Therefore we introduce a general rule for our views instead.

1.3.1 Data Sets: General Description

We have two different data sets for the implementation of our model: The MSCI World index and the S&P EMBI. The MSCI World index data set includes daily returns in USD for the period February 1994 to May 2014. The number of assets for this data set is $n = 940$. The S&P EMBI data set includes daily returns in USD for the period January 1996 to July 2014. The number of assets for this data set is $n = 591$.

1.3.2 Back-Testing Approach

Table 1.3: Back-Testing Approach.

	S&P 500 EMBI	MSCI
Number of available assets n	591	940
First available data for Σ	January 1996	January 1995
First adjustment back-testing	January 2006	January 2005
Final adjustment back-testing	October 2013	October 2013
Frequency of adjustments	Quarterly	Quarterly
Number of adjustments A	32	36

This table provides the dates for our back-testing approach.

We assess the quality of our model with a back-testing approach. An overview of the dates for this approach is provided in Table 1.3. In a first step we calculate the sample covariance matrix Σ using the daily returns from 10 years, more precisely from January 1996 to December 2005 for the S&P EMBI and from January 1995 to December 2004 for the MSCI World index. Then we run the GMV optimization model at January 2006 for the S&P EMBI (and at January 2005 for the MSCI World index) to find the GMV portfolio. Then we adjust the GMV portfolio with our modified B–L

[4]I thank OLZ for providing the data sets. I am especially grateful to Lorenz Beyeler, Cyril Bachelard, and Carmine Orlacchio for professional advice and helpful discussions regarding the topic.

approach.

We readjust our portfolio every 90 days, which is similar to the practice at OLZ where portfolios are adjusted quarterly. The variance-covariance matrix is always based on all the data from the first available data point in time (January 1996 for the S&P EMBI and January 1995 for the MSCI World index) until the time of the readjustment. We choose this approach instead of a moving horizon, which only uses the data of the past 10 years at every adjustment point, to provide as much available data as possible for our optimization model. In total we optimize and adjust the portfolio 32 times (8 years, 4 times a year) for the S&P EMBI and 36 times (9 years, 4 times a year) for the MSCI World index, with the last optimization and adjustment being carried out in October 2013.

To assess the quality of our model we calculate the out-of-sample returns for the three months following each optimization and adjustment of the portfolio. We then multiply the out-of-sample returns with the weights of our portfolio. We compare the cumulative out-of-sample returns (CRs) of our model to the CRs of the GMV portfolio and the $1/N$ portfolio.

1.3.3 A General Rule to Simulate Investors' Views

Since we are assessing our model's quality with a back-testing approach it is complicated to simulate the views for the parameters P and q. Instead of creating certain single views, which could be biased because we already know the events of the future, we use a general rule for the view parameters instead. In real life this general rule could still be supplemented with investors' specific views.

The general rule is based on the expected returns for each asset and the systematic risk of each asset. We calculate the expected returns $\bar{R}_j$ of the j^{th} asset using the sample mean formula

$$\bar{R}_j = \frac{1}{S}\sum_{s=1}^{S} R_{s,j} \quad \forall\ j. \tag{1.14}$$

The systematic risk β_j of the jth asset is calculated with the following formula (see Black et al. (1972)):

$$\beta_j = \frac{cov(\bar{R}_j, \bar{R}_M)}{\sigma^2(\bar{R}_M)}, \tag{1.15}$$

where $\bar{R}_M$ is the expected average return of all n candidate assets. In a next step we introduce the parameter v, which can be between 0 and n. If the expected return $\bar{R}_j$ and the β_j of an asset are both among the v smallest values of all assets n, we create a view for this asset. In this way we get $k \leq v \leq n$ views. Thereby, we only analyze absolute views in this paper.

To express which assets are involved in the view we use the matrix P. This matrix contains one row k for each view. Each of these rows has one value set to 1 at the index j that we have the view on. All the other values are set to zero.

For both data sets we run our model for eight different values of v. To ensure comparability of the results for both data sets we set the values of v as percentages of the total available assets n. For more details see Table 1.4.

The actual size of the views is expressed with the column vector q consisting of k rows. To keep the model simple and to avoid the potential benefits of knowing the future while carrying out the back-testing we do not determine each value in q individually. Instead we set all elements in this vector to a value that is considerably smaller than the corresponding values in the excess return vector Π_N (specifically we choose a value of 0.0001). This should lead to considerably smaller holdings in the assets we have views on compared to the GMV model. In Section 1.4.3.2 we analyze the effect of a deviation from the value 0.0001.

Table 1.4: Values for Our General Rule.

	S&P 500 EMBI	MSCI
$v = 0.25n$	148	235
$v = 0.3n$	177	282
$v = 0.35n$	207	329
$v = 0.4n$	236	376
$v = 0.45n$	266	423
$v = 0.5n$	296	470
$v = 0.55n$	325	517
$v = 0.6n$	355	564

This table provides the general rule parameters for our back-testing approach. If v was not a natural number we rounded it to the next natural number.

1.3.4 Confidence of 100 Percent in the Views

In Section 1.3.3 we introduced a general rule for our views. In the context of the generality of this rule, it would be arbitrary to determine the confidence of our views using the parameter Ω. Instead we assume 100 percent confidence in the views created by the general rule. This enables us to set all the diagonal elements of Ω to zero. According to Idzorek (2005), we can therefore erase Ω from Equation (1.3). This leads to the following formula for calculating our new combined return vector μ^*:

$$\mu^* = \Pi_N + \tau\Sigma P'(\tau P\Sigma P')^{-1}(q - P\Pi_N). \tag{1.16}$$

Walters (2014) emphasizes that the combined return vector of Equation (1.16) is insensitive to the value of τ. So we can erase the weight-on-views from Equation (1.16), which leads to the final formula

[5] There is high uncertainty in the literature with regards to the best value for τ. Walters (2013) even dedicates a whole paper to the factor of τ as one of the more confusing aspects of the model. Allaj (2013) identifies the calibration of τ as one of the two well-known problems of the B–L model and suggests a precise formula for τ by using an econometric model. Black and Litterman (1990) suggest a weight-on-views parameter value close to zero while Satchell and Scowcroft (2000) use a value of $\tau = 1$. Bevan and Winkelmann (1998) argue that the value used in practice lies somewhere in the middle with a τ between 0.5 and 0.7. Idzorek (2005) states that the easiest way to calibrate the B–L model is to make an assumption about the weight-on-views and suggests $\tau = 0.025$.

for calculating μ^*:

$$\mu^* = \Pi_N + \Sigma P'(P\Sigma P')^{-1}(q - P\Pi_N). \tag{1.17}$$

The absence of the weight-on-views parameter τ is a beneficial side effect of the assumption of 100 percent confidence in our views, since it is known to be the most critical parameter in the B–L model.[5]

1.3.5 Specific Problem Instance

Table 1.5: Parameter Values.

Symbol	Value	Description	Size
Σ	-	Estimated covariance matrix	n x n
δ	3.07	Risk aversion parameter	1 x 1
P	-	Assets involved in the views	k x n
q	0.0001 (each element)	View vector	1 x k

This table provides specific parameter values for the model's implementation.

For the first step of our model, the optimization of the GMV portfolio, we have to calculate estimators for the variance-covariance matrix Σ based on daily returns

$$\Sigma_{i,j} = 1/(S-1)\sum_{s=1}^{S}(R_{s,i} - \bar{R}_i)(R_{s,j} - \bar{R}_j), \tag{1.18}$$

where $R_{s,i}$ are historical daily returns. For the B–L approach we have to determine the values of certain parameters. We use the same estimated variance-covariance matrix that we use for the optimization of the GMV portfolio. It is a square matrix with n rows and n columns. Suggestions for the risk aversion parameter δ differ in the B–L literature. He and Litterman (1999) use "$\delta = 2.5$ as the risk parameter representing the world average risk tolerance". Idzorek (2005) relates the risk aversion to the risk premium and the variance of the market excess returns and determines $\delta = 3.07$. We decided to go with the latter value. As mentioned in Section 1.3.3 we use the matrix P to express which assets are involved in the views. This matrix consist of k rows, one for each view. It has n columns; $n-1$ elements per row have a value of zero; the element j which we have a view on, has a value of 1. We also mentioned in Section 1.3.3 that we use a value of 0.0001 for all k elements of the view vector q to ensure smaller holdings in the assets we have views on for our model compared to the GMV portfolio. Table 1.5 provides an overview of the specific parameter values for our model's implementation.

1.4 Results

This results section consists of four subsections. In the first, we analyze the out-of-sample returns of eight B–L portfolios with different values of v and the GMV portfolio and the $1/N$ portfolio, which both serve as benchmarks. In the second subsection we depict and analyze the compound annual returns ($CARs$) and the annual standard deviation (σ_{an}) for all these portfolios. Furthermore, we show the out-of-sample Sharpe ratios of the eight B–L portfolios, the GMV portfolio, and the $1/N$ portfolio benchmarked against the risk-free rate. We also provide the p-values of the difference between the Sharpe ratio of each of the B–L portfolios and that of the GMV portfolio and the $1/N$ portfolio. In the third subsection we evaluate the effect of the variation of certain parameters of the model. Finally, in the fourth subsection we show the composition and the level of diversification of the eight analyzed B–L portfolios.

1.4.1 Out-of-Sample Portfolio Returns

In a first step to assess the quality of our modified B–L model we analyze the out-of-sample quarterly portfolio returns from 2006 until 2013 for the S&P EMBI data set and from 2005 until 2013 for the MSCI data set. We benchmark these portfolio returns against the out-of-sample portfolio returns of the GMV portfolio and the $1/N$ portfolio. We use the GMV portfolio as a benchmark since it is widely represented in practice, including at OLZ. Additionally we use the $1/N$ portfolio since it sometimes outperforms mean-variance portfolios in practice (see De Miguel et al. (2009)). However, in contrast to De Miguel et al. (2009) we do not consider transaction costs.

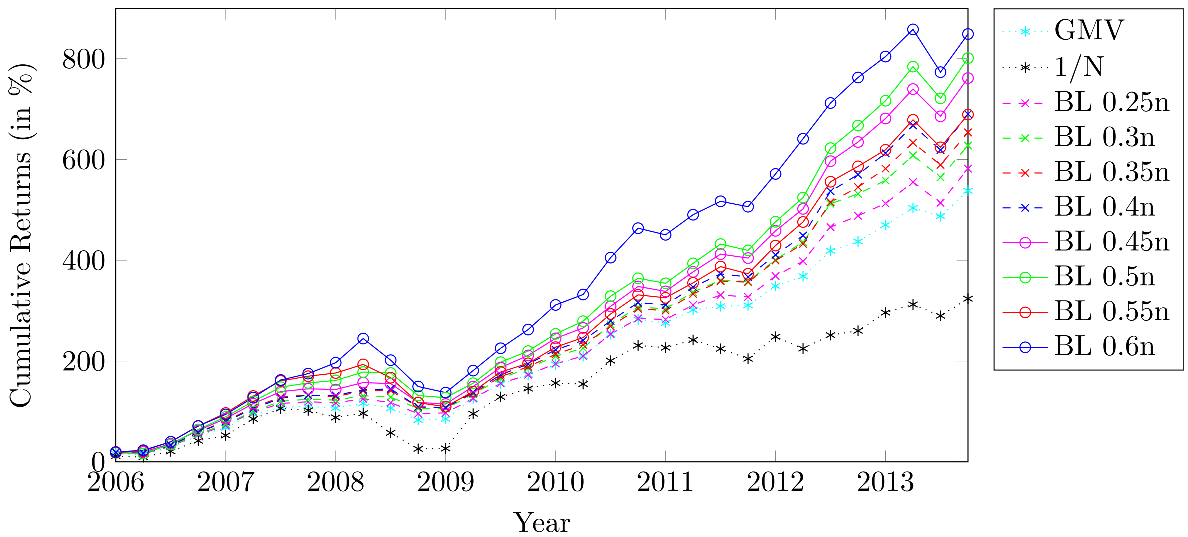
1.4.1.1 Cumulative Out-of-Sample Returns at the Adjustment Points

To calculate the cumulative returns of our portfolios we first add 1 to the quarterly returns and multiply the resulting factors together. Secondly we subtract 1 from the final result of the multiplication to get the cumulative returns at the end of the period

$$CR = \prod_{a=1}^{A} (r_a + 1) - 1. \tag{1.19}$$

Figure 1.1 shows the CRs of the eight modified B–L portfolios, the GMV portfolio, and the $1/N$ portfolio for the S&P EMBI data set from 2006 until 2013. We can see that the out-of-sample returns of all of the eight portfolios develop in a relatively similar manner and achieve very high CRs over time. We also see that all eight modified B–L portfolios outperform the GMV portfolio and the $1/N$ portfolio in terms of out-of-sample returns. The modified B–L portfolio with $v = 0.25n$ only has a slightly larger CR by the end of 2013 than the GMV portfolio (581.7% compared to 538.0%). However, once we increase the value of v, the CRs of the modified B–L portfolios increase and the modified portfolio with $v = 0.6$ has a CR of 849.0% by the end of 2013, which is more than 1.5 times as high as the CR of the GMV portfolio. An exception is the B–L portfolio with $v = 0.55$. This portfolio only has a CR of 689.1%, which is smaller than the CRs of our B–L portfolio with a

Figure 1.1: Cumulative Out-of-Sample Returns (S&P EMBI).

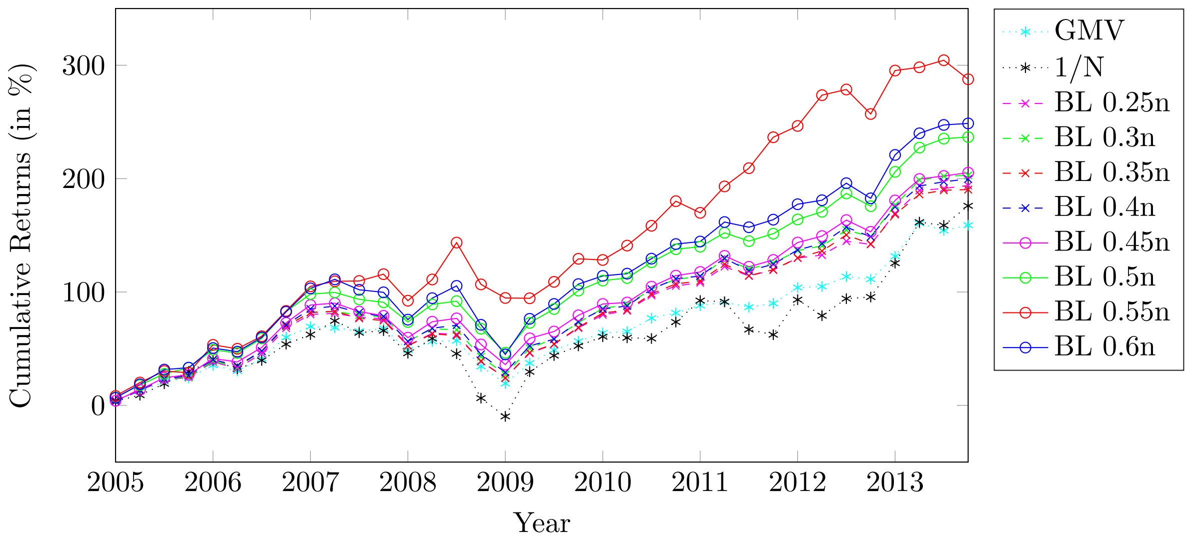


This figure shows the cumulative out-of-sample quarterly returns of eight B–L portfolios, the GMV portfolio, and the $1/N$ portfolio from 2006 until 2013. The portfolios are readjusted quarterly (S&P EMBI).

v value of $0.4n$, but still clearly larger than the CR of the GMV portfolio. The naive diversification approach with a $1/N$ portfolio results in clearly smaller CRs than the B–L or the GMV approach.

Figure 1.2 shows the CRs of eight modified B–L portfolios, the GMV portfolio, and the $1/N$ portfolio for the MSCI data set from 2005 until 2013. In general the out-of-sample returns of the B–L portfolios and the GMV portfolio in this data set develop in a similar manner. The out-of-sample returns of the $1/N$ portfolio show a slightly different pattern than the out-of-sample returns of the GMV portfolio and the B–L portfolios. At some adjustment points the magnitude of the out-of-sample returns of the $1/N$ portfolio is much larger than that of the other nine portfolios. For example, in the fourth quarter of 2008 the CR of the $1/N$ portfolio decreases much stronger than the CR of the other portfolios. Then again in the second quarter of 2009 the CR of the $1/N$ portfolio increases much more strongly than those of the other portfolios. At certain adjustment points, the direction of the out-of-sample returns of the $1/N$ portfolio also differ, such as in the second quarter of 2010 with -0.6% compared to returns between 0.7% and 5.6%. By the end of 2013 all eight modified B–L portfolios have outperformed both the GMV portfolio and the $1/N$ portfolio considerably. In contrast to the S&P EMBI data set, the modified B–L portfolio with $v = 0.55$ has the highest CRs for the MSCI data set, while the portfolio with $v = 0.6$ only has the second highest CR. All the other modified B–L portfolios show quite a similar performance and there is no constant improvement for an increase of v, which could—with the exception of the B–L portfolio with $v = 0.55$—be seen for the S&P EMBI data set. The $1/N$ portfolio has a larger CR than the GMV portfolio, which is in line with the findings of De Miguel et al. (2009). In general all portfolios created with the MSCI data set do not perform as well as the portfolios of the S&P EMBI data set, which can be explained

Figure 1.2: Cumulative Out-of-Sample Returns (MSCI).



This figure shows the cumulative out-of-sample quarterly returns of eight B–L portfolios, the GMV portfolio, and the $1/N$ portfolio from 2005 until 2013. The portfolios are readjusted quarterly (MSCI).

by the nature of the data (world index vs. emerging markets).

1.4.1.2 Out-of-sample Returns at the Adjustment Points

In the next step of our analysis we take a closer look at the out-of-sample returns over time. Therefore we compare the out-of-sample returns at every adjustment point of the B–L portfolio with $v = 0.5n$ and the GMV portfolio. We showed in Section 1.4.1.1 that all eight B–L portfolios behave in a relatively similar manner in a cumulative sense. This pattern is also apparent for the quarterly returns. Therefore we limit ourselves to just one of the eight B–L portfolios and choose the $v = 0.5n$ portfolio. This has the second highest CR of all portfolios by the end of 2013 for the S&P EMBI data set and the third highest CR by the end of 2013 for the MSCI data set. The only portfolio with a larger CR by the end of 2013 for both data sets is the B–L portfolio with $v = 0.6$. However, we can see from Table 1.6 that the latter has a much larger annual standard deviation than the portfolio with $v = 0.5$ (We discuss annual standard deviations in more detail in Section 1.4.2). Therefore we consider the portfolio with $v = 0.5$ as a more suitable representative of the B–L portfolios. As a benchmark for our comparison we use the GMV portfolio.

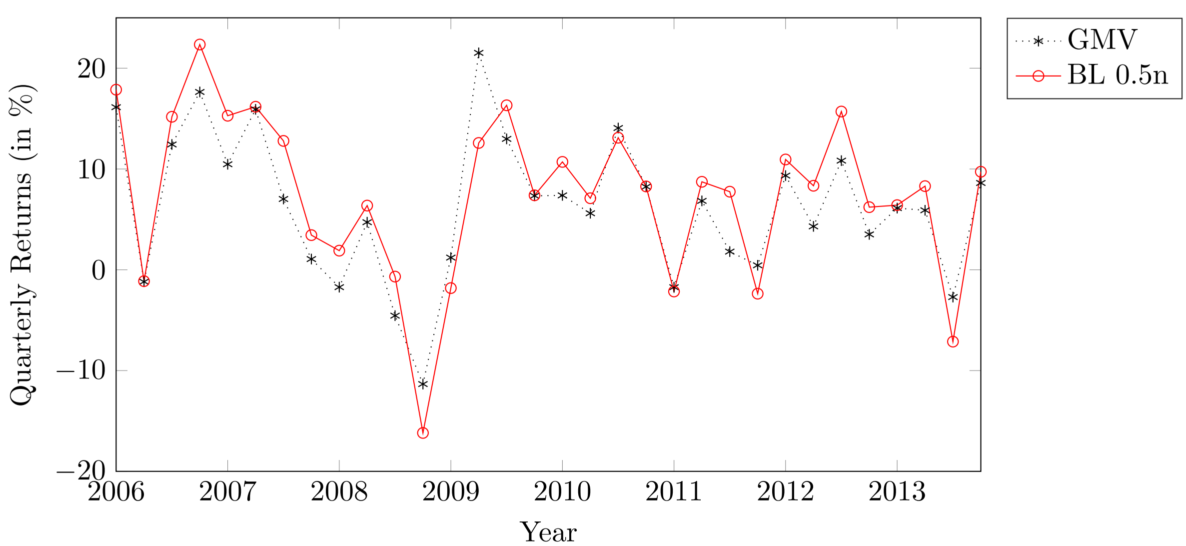
Figure 1.3 shows the out-of-sample returns at every adjustment point of the B–L portfolio with $v = 0.5n$ and the GMV portfolio for the S&P EMBI data set from 2006 until 2013. In general the out-of-sample returns of both portfolios are relatively similar and move in the same direction at every adjustment point. However, the B–L portfolio outperforms the GMV portfolio at 25 out of 32 adjustment points and at some of them significantly. The B–L portfolio only performs considerably worse than the GMV portfolio in the second quarter of 2009 (the out-of-sample return of the B–L

Table 1.6: Cumulative Out-of-Sample Returns by the End of 2013, Compound Annual Returns, and Annual Standard Deviations.

	S&P EMBI			MSCI		
Portfolio	CR	CAR	σ_{an}	CR	CAR	σ_{an}
$v = 0.25n$	581.7%	27.1%	14.1%	193.9%	12.7%	14.0%
$v = 0.3n$	627.7%	28.2%	13.9%	202.8%	13.1%	14.0%
$v = 0.35n$	653.6%	28.7%	14.5%	190.3%	12.6%	14.1%
$v = 0.4n$	690.1%	29.5%	14.9%	199.6%	13.0%	14.3%
$v = 0.45n$	761.5%	30.9%	15.5%	205.2%	13.2%	13.7%
$v = 0.5n$	801.2%	31.6%	16.2%	236.6%	14.4%	13.6%
$v = 0.55n$	689.1%	29.5%	17.4%	270.7%	17.8%	15.1%
$v = 0.6n$	849.0%	32.5%	18.7%	248.7%	16.9%	16.0%
GMV	538.0%	26.1%	14.2%	158.9%	11.1%	13.0%
$1/N$	323.7%	19.8%	26.7%	176.0%	11.9%	23.5%

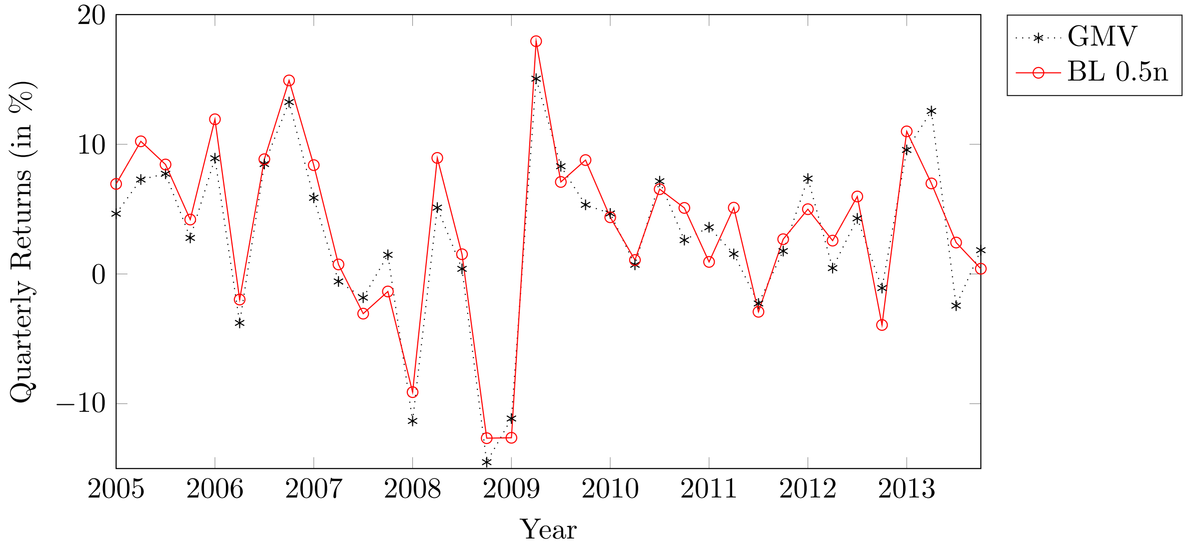
This table reports the cumulative out-of-sample returns by the end of 2013 (CR), compound annual returns (CAR), and annual standard deviations (σ_{an}) for eight B–L portfolios, the GMV portfolio, and the $1/N$ portfolio.

Figure 1.3: Quarterly Out-of-Sample Returns (S&P EMBI).



This figure shows the out-of-sample quarterly returns of the GMV portfolio and the B–L portfolio with $v = 0.5n$ at every adjustment point (S&P EMBI).

Figure 1.4: Quarterly Out-of-Sample Returns (MSCI).



This figure shows the out-of-sample quarterly returns of the GMV portfolio and the B–L portfolio with $v = 0.5n$ at every adjustment point (MSCI).

portfolio is 12.6% compared to an equivalent figure of 21.5% for the GMV portfolio).

Figure 1.4 shows the out-of-sample returns at every adjustment point of the B–L portfolio with $v = 0.5n$ and the GMV portfolio for the MSCI data set from 2005 until 2013. In general the out-of-sample returns of both portfolios have the same direction and the same magnitude. However the B–L portfolio has larger out-of-sample returns at 24 out of 36 adjustment points while the GMV portfolio only performs better at 12 adjustment points. The B–L portfolio only performs considerably worse than the GMV portfolio in the second quarter of 2013 (the out-of-sample returns are 7.0% and 12.6%, respectively).

1.4.2 Compound Annual Returns, Standard Deviations, and Sharpe Ratios

The analysis of the out-of-sample returns in Section 1.4.1 has already provided a good indication of the performance of our modified B–L model. However, it is not sufficient for an extensive assessment. Therefore we consider additional performance indicators in this section. First, the compound annual return (CAR) and the annual standard deviation (σ_{an}). Second, the out-of-sample Sharpe ratio. As in the previous subsection we use the GMV portfolio and the $1/N$ portfolio as benchmarks.

1.4.2.1 Compound Annual Returns and Standard Deviations

To calculate the CAR we add 1 to the cumulative return. Then we take the $A/4th$ root (A equals the number of quarters, so $A/4$ equals the number of years) of that sum. Finally we subtract 1 from that result to determine the CAR

$$CAR = \sqrt[A/4]{(CR+1)} - 1. \tag{1.20}$$

We also calculate the standard deviation σ_q of the quarterly out-of-sample returns. In a next step we annualize those standard deviations using the standard method of the industry. The following formula is used to calculate σ_{an}:

$$\sigma_{an} = \sigma_q * \sqrt{4}. \tag{1.21}$$

In the second and third column beneath Table 1.6's "S&P EMBI" heading, we show the CAR and σ_{an} of the ten portfolios for the S&P EMBI data set. The CAR of all eight modified B–L portfolios is larger than the $CARs$ of the GMV portfolio and the $1/N$ portfolio. As for the cumulative returns, the $CARs$ of the B–L portfolios increases with an increase of v (with the exception of the portfolio with $v = 0.55$). Furthermore, the $1/N$ portfolio has a clearly smaller CAR value than all the other portfolios. The values of σ_{an} for most of the B–L portfolios are a little larger than that for the GMV portfolio. Among the B–L portfolios the highest CAR, of the portfolio with $v = 0.6$, comes with the price of the highest σ_{an}, of 18.7 percent. The $1/N$ portfolio not only has by far the smallest CAR of all ten portfolios, it also has a σ_{an} of 26.7 percent, which is considerably larger than the σ_{an} of the other nine portfolios.

In the second and third column beneath Table 1.6's "MSCI" heading, we show the CAR and σ_{an} of the ten portfolios for the MSCI data set. The modified B–L portfolios have a larger CAR than those of both the GMV and the $1/N$ portfolios. The CAR of the B–L portfolios in the MSCI data set does not constantly increase with an increase of v. We can see that the CARs of the five B–L portfolios with values of v between 0.25 and 0.45 are at a similar level (between 12.6% and 13.2%), while the $CARs$ of the three other B–L portfolios are clearly larger (between 14.4% and 17.8%). The $1/N$ portfolio has a larger CAR than the GMV portfolio. All B–L portfolios and the GMV portfolio have a rather similar σ_{an} (between 13.0% and 16.0%). The σ_{an} value for the $1/N$ portfolio is considerably larger (23.5%).

1.4.2.2 Out-of-Sample Sharpe Ratios

Our next performance indicator is the out-of-sample Sharpe ratio. Following Sharpe (1994), we compare our eight modified B–L portfolios, the GMV portfolio, and the $1/N$ portfolio to a benchmark. The risk-free rate serves as our benchmark.[6] At each adjustment point we subtract the risk-free rate from the out-of-sample return of every modified B–L portfolio, the GMV portfolio, and the $1/N$ portfolio. This provides us the excess returns D of the B–L portfolios, the GMV portfolio, and the $1/N$ portfolio R_P over the risk-free rate benchmark R_F

$$D = R_{P,t} - R_{F,t} \quad \forall\ t. \tag{1.22}$$

[6] We use the risk-free rate of the Fama/French Global Factors; see `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/details_global.html` for further information (last accessed March 21, 2016).

Table 1.7: Sharpe Ratios.

	S&P EMBI			MSCI		
Portfolio	SR	p-value vs.		SR	p-value vs.	
		GMV	$1/N$		GMV	$1/N$
$\upsilon = 0.25n$	0.8762	0.25	0.00	0.4141	0.30	0.09
$\upsilon = 0.3n$	0.9176	0.13	0.00	0.4261	0.23	0.07
$\upsilon = 0.35n$	0.8998	0.21	0.00	0.4057	0.37	0.10
$\upsilon = 0.4n$	0.8987	0.22	0.00	0.4149	0.32	0.09
$\upsilon = 0.45n$	0.9057	0.18	0.00	0.4392	0.20	0.06
$\upsilon = 0.5n$	0.8913	0.28	0.00	0.4834	0.06	0.03
$\upsilon = 0.55n$	0.7816	0.29	0.00	0.4802	0.10	0.03
$\upsilon = 0.6n$	0.8014	0.39	0.00	0.4355	0.25	0.07
GMV	0.8337	-	-	0.3852	-	-
$1/N$	0.3782	-	-	0.2645	-	-

This table reports the quarterly out-of-sample Sharpe ratios (SR) for the eight B–L portfolios, the GMV portfolio, and the $1/N$ portfolio benchmarked against the risk-free rate. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the eight B–L portfolios and the Sharpe ratios of the GMV portfolio and the $1/N$ portfolio, respectively.

Then we calculate the mean and the standard deviation of the excess returns D. Finally we divide the mean by the standard deviation for each of these ten portfolios to find the out-of-sample Sharpe ratio SR

$$SR = \frac{\bar{D}}{\sigma_D}. \tag{1.23}$$

We also test whether the Sharpe ratios of our B–L portfolios are statistically distinguishable from the Sharpe ratio of the GMV portfolio and the Sharpe ratio of the $1/N$ portfolio. As do De Miguel et al. (2009), we therefore compute the p-value of the difference, using the Jobson and Korkie (1981) approach and the correction of Memmel (2003).[7]

In the three columns under the heading “S&P EMBI” in Table 1.7 we show the out-of-sample Sharpe ratios for the S&P EMBI data set and the p-values of the difference between the out-of-sample Sharpe ratios of the eight B–L portfolios and both the GMV portfolio and the $1/N$ portfolio. We

[7] For our out-of-sample returns of the two portfolios i and n we obtain the test of the hypothesis $H_0 : \hat{\mu}_i/\hat{\sigma}_i - \hat{\mu}_n/\hat{\sigma}_n = 0$ using the test statistic $\hat{z}_{JK}$, which is asymptotically distributed as a standard normal

$$\hat{z}_{JK} = \frac{\hat{\sigma}_n\hat{\mu}_i - \hat{\sigma}_i\hat{\mu}_n}{\sqrt{\hat{\vartheta}}},\ with\ \hat{\vartheta} = \frac{1}{A}\left(2\hat{\sigma}_i^2\hat{\sigma}_n^2 - 2\hat{\sigma}_i\hat{\sigma}_n\hat{\sigma}_{i,n} + \frac{1}{2}\hat{\mu}_i^2\hat{\sigma}_n^2 + \frac{1}{2}\hat{\mu}_n^2\hat{\sigma}_i^2 - \frac{\hat{\mu}_i\hat{\mu}_n}{\hat{\sigma}_i\hat{\sigma}_n}\hat{\sigma}_{i,n}^2\right), \tag{1.24}$$

where $\hat{\mu}_i$, $\hat{\mu}_n$, $\hat{\sigma}_i$, $\hat{\sigma}_n$, and $\hat{\sigma}_{i,n}$ are the estimated means, variances, and covariances and A is the number of adjustments (see De Miguel et al. (2009)).

first notice that each B–L portfolio has a statistically significant larger Sharpe ratio at the 0.01 level than the $1/N$ portfolio, which has a Sharpe ratio of 0.3782 (the p-values for all eight portfolios are 0.00). The B–L portfolios between $v = 0.25$ and $v = 0.5$ have Sharpe ratios between 0.8762 and 0.9176, which are all larger than the Sharpe ratio of 0.8337 for the GMV portfolio. However with p-values between 0.13 and 0.28 this difference is not statistically significant at conventional levels. The two B–L portfolios with $v = 0.55$ and $v = 0.6$ have slightly smaller Sharpe ratios than the GMV portfolio. However, this difference is also not statistically significant.

In the three columns under the heading "MSCI" in Table 1.7 we show the out-of-sample Sharpe ratios for the MSCI data set and the p-values of the difference between the out-of-sample Sharpe ratios of the eight B–L portfolios and both the GMV portfolio and the $1/N$ portfolio. As for the S&P EMBI data set, each B–L portfolio has a statistically significant larger Sharpe ratio than the $1/N$ portfolio, which has a Sharpe ratio of 0.2645. For the B–L portfolios with $v = 0.5$ and $v = 0.55$ the difference is statistically significant at the 0.05 level (the p-value is 0.03 for both portfolios). For the other six portfolios, with p-values between 0.06 and 0.10, the difference is only significant at the 0.10 level. The eight B–L portfolios have Sharpe ratios between 0.4057 and 0.4834, which are all larger than the Sharpe ratio of 0.3852 for the GMV portfolio. For the portfolios with $v = 0.5$ and $v = 0.55$ the difference is statistically significant at the 0.10 level (p-values are 0.06 and 0.10, respectively). The Sharpe ratios of the other six B–L portfolios have differences that are statistically insignificant (p-values are between 0.20 and 0.37).

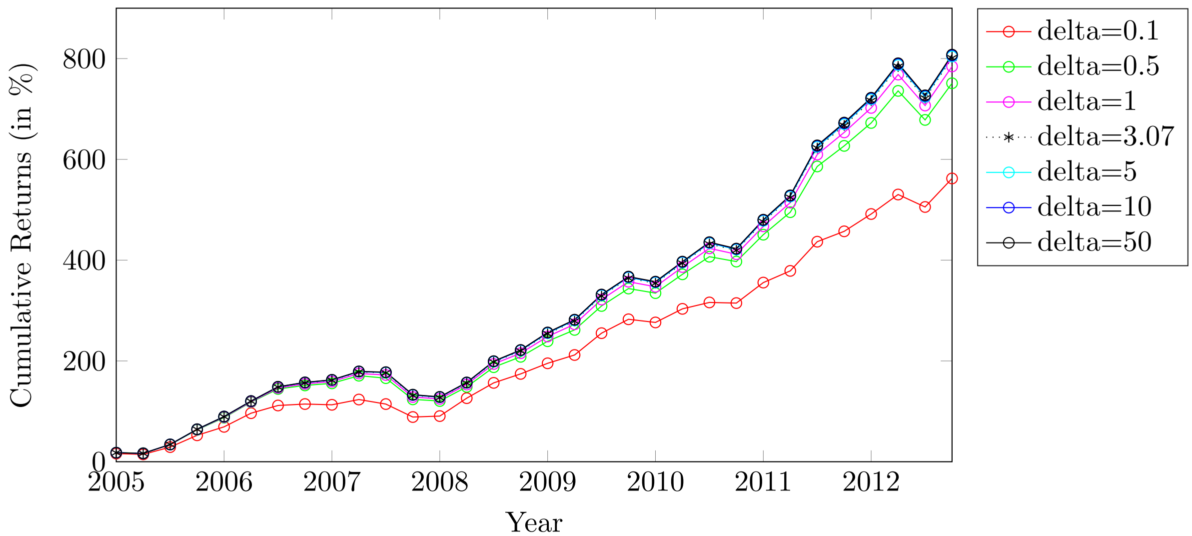
1.4.3 Variations of Other Parameters

After analyzing the performance of several B–L portfolios with different values of v in the previous sections we now focus on the sensitivity of the other parameters of our model. We consider three parameters: the risk aversion parameter δ, the elements of the view vector q, and the time parameter. In addition we assess the effect of different time periods on the CR and the Sharpe ratio. For the sensitivity analysis we focus on the B–L portfolio with $v = 0.5$ because it has the highest Sharpe ratio for the MSCI data set and one of the highest for the S&P EMBI data set. As benchmarks we use the GMV portfolio and the $1/N$ portfolio. We discuss the parameter sensitivities for both data sets to analyze similarities and differences.

1.4.3.1 Variations of δ

Figure 1.5 shows the CRs of the B–L portfolio with $v = 0.5$ for seven different values of δ for the S&P EMBI data set. Besides the original δ value of 3.07 we used values of 0.1, 0.5, 1, 5, 10, and 50 for δ. According to Idzorek (2005) a larger δ in the original B–L setup means more excess return per unit of risk and therefore an increase in the estimated excess returns. We can see from Table 1.8 that an increase in δ leads to an increase in the CR in 2013 and the $CARs$ for the portfolio. However, an increase in δ also leads to an increase in σ_{an} for the portfolio. The increase in CRs in 2013 and $CARs$ between the portfolios with a δ of 1 and 50 is relatively small (from 784.4% to 807.8%). However the out-of-sample returns and the $CARs$ decrease considerably for values of δ smaller than 1. For

Figure 1.5: Cumulative Out-of-Sample Returns of the BL 0.5n Portfolio for Different δ (S&P EMBI).



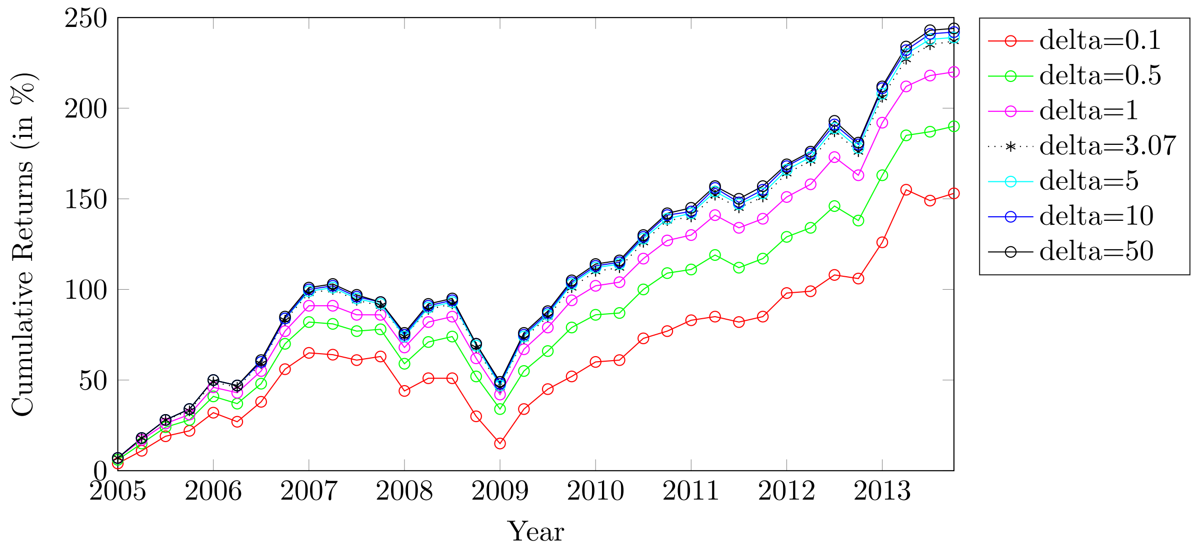
This figure shows the cumulative out-of-sample quarterly returns for the v=0.5n B–L portfolio for seven different values of δ from 2006 until 2013. The portfolios are readjusted quarterly (S&P EMBI).

Table 1.8: Variations of δ (S&P EMBI).

δ	CR	CAR	σ_{an}	SR	p-value vs.	
					GMV	$1/N$
0.1	562.2%	26.7%	14.1%	0.8575	0.18	0.00
0.5	751.0%	30.7%	15.9%	0.8834	0.29	0.00
1	784.4%	31.3%	16.0%	0.8909	0.28	0.00
3.07	801.2%	31.6%	16.2%	0.8913	0.28	0.00
5	804.1%	31.7%	16.2%	0.8923	0.28	0.00
10	806.3%	31.7%	16.2%	0.8932	0.27	0.00
50	807.8%	31.7%	16.2%	0.8938	0.27	0.00

This table reports the cumulative out-of-sample returns by the end of 2013 (CR), compound annual returns (CAR), annual standard deviation (σ_{an}), and the out-of-sample Sharpe ratio for the v=0.5n B–L portfolio for seven different values of δ. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the Sharpe ratios of the GMV portfolio (0.8337) and the $1/N$ portfolio (0.3782), respectively (S&P EMBI).

Figure 1.6: Cumulative Out-of-Sample Returns of the B–L 0.5n Portfolio for Different δ (MSCI).



This figure shows the cumulative out-of-sample quarterly returns for the v=0.5n B–L portfolio for seven different values of δ from 2005 until 2013. The portfolios are readjusted quarterly (MSCI).

values of δ close to zero, the B–L portfolio has similar values to the GMV portfolio, not only in terms of out-of-sample returns and $CARs$, but also in terms of the portfolio composition. The B–L portfolio with $\delta = 0.1$ has a CR in 2013 of 562.2 percent, which is relatively similar to the CR of the GMV portfolio of 538 percent. For all values of δ the B–L portfolio has a larger Sharpe ratio than the GMV portfolio, which has a ratio of 0.8337. However, the difference is statistically insignificant for all values of δ (p-values are between 0.18 and 0.29). For all values of δ the B–L portfolio has a statistically significant larger Sharpe ratio at the 0.01 level than the $1/N$ portfolio, which has a Sharpe ratio of 0.3782 (the p-values for all eight portfolios are 0.00).

Figure 1.6 shows the CRs for the B–L portfolio with $v = 0.5$ for seven different values of δ for the MSCI data set. Besides the original δ value of 3.07 we used values of 0.1, 0.5, 1, 5, 10, and 50 for δ. The results are very similar to those for the S&P EMBI data set. We can see from Table 1.9 that an increase of δ leads to an increase of the CR in 2013 and the $CARs$. The values of σ_{an} are similar for all seven portfolios, but have the tendency to increase slightly when δ increases. As for the S&P EMBI data set, for our initial value of $\delta = 3.07$ an increase of δ has a relatively small effect. However, for δ values smaller than 1, a decrease of δ has a clear negative effect. For all values of δ besides 0.1 the Sharpe ratio of the B–L portfolio is larger than that of the GMV portfolio, which has a Sharpe ratio of 0.3852. However, the difference is only statistically significant at the 0.10 level for δ values of 1 and larger. The Sharpe ratio of the B–L portfolio is larger than that of the $1/N$ portfolio for all values of δ. However, the difference is only statistically significant at the 0.05 level for δ values of 0.5 and larger.

Table 1.9: Variations of δ (MSCI).

δ	CR	CAR	σ_{an}	SR	p-value vs.	
					GMV	$1/N$
0.1	153.3%	10.9%	13.1%	0.3713	0.10	0.13
0.5	190.1%	12.6%	12.9%	0.4375	0.12	0.05
1	219.9%	13.8%	13.3%	0.4704	0.07	0.03
3.07	236.6%	14.4%	13.6%	0.4834	0.06	0.03
5	239.4%	14.5%	13.6%	0.4853	0.06	0.03
10	242.0%	14.6%	13.7%	0.4874	0.06	0.03
50	243.9%	14.7%	13.7%	0.4890	0.06	0.03

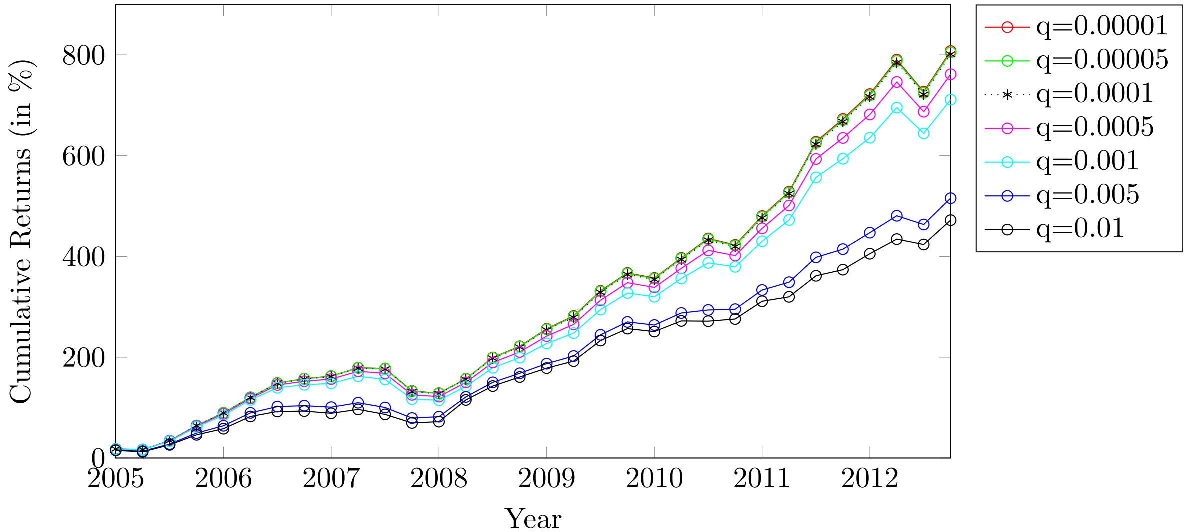
This table reports the cumulative out-of-sample returns by the end of 2013 (CR), compound annual returns (CAR), annual standard deviation (σ_{an}), and the out-of-sample Sharpe ratio for the v=0.5n B–L portfolio for seven different values of δ. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the Sharpe ratios of the GMV portfolio (0.3852) and the $1/N$ portfolio (0.2645), respectively (MSCI).

1.4.3.2 Variations of the Element Values of q

Figure 1.7 shows the CRs for the B–L portfolio with $v = 0.5n$ for seven different values for the elements of q for the S&P EMBI data set. Besides the original value for the elements of q of 0.0001 we used values of 0.00001, 0.00005, 0.0005, 0.001, 0.005, and 0.01 for the elements of q. We can see that the value of the elements of q has a strong influence on the CR in 2013. A reduction of the q value leads to a larger CR. Table 1.10 shows that the larger cumulative returns come with larger $CARs$. However, the effect of the q values on σ_{an} is ambiguous. The value of σ_{an} decreases from 16.2 percent for a value of 0.00001 for all elements of q to 14.0 percent at q element values of 0.005. A further reduction of the elements of q to 0.01 leads to an increase of σ_{an} to 16.2 percent. The values of the elements of q also have a strong influence on the Sharpe ratios of the B–L portfolio. For values for the elements of q between 0.00001 and 0.001 the Sharpe ratios are at a similar level and larger than the Sharpe ratio of the GMV portfolio. However the difference is statistically insignificant. For the two largest values for the elements of q the Sharpe ratio of the B–L portfolio is smaller than that of the GMV portfolio. For all chosen q values the B–L portfolio has a statistically significant larger Sharpe ratio than the $1/N$ portfolio at the 0.01 level.

Figure 1.8 shows the CRs of the B–L portfolio with $v = 0.5$ for seven different values for the elements of q for the MSCI data set. As for the S&P EMBI data set we used the original value for the elements of q of 0.0001 and six additional values of 0.00001, 0.00005, 0.0005, 0.001, 0.005, and 0.01 for the elements of q. The value of the elements of q has a strong effect on the CR of the portfolios. The CR in 2013 increases considerably with a reduction of the values of the elements of

Figure 1.7: Cumulative Out-of-Sample Returns of the B–L 0.5n Portfolio for Different q (S&P EMBI).



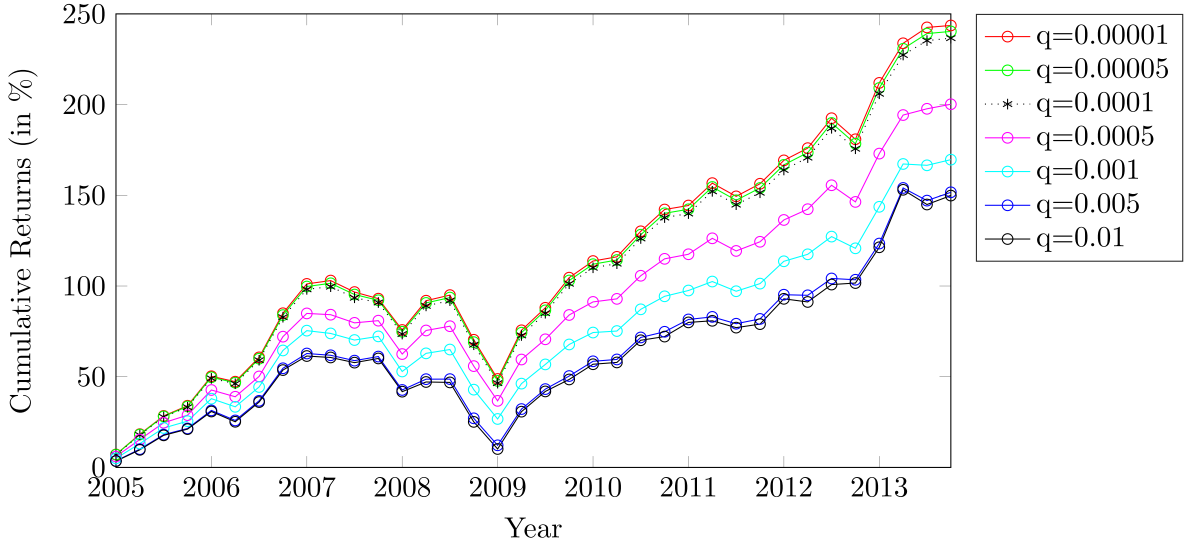
This figure shows the cumulative out-of-sample quarterly returns for the v=0.5n B–L portfolio for seven different values of q from 2006 until 2013. The portfolios are readjusted quarterly (S&P EMBI).

Table 1.10: Variation of q (S&P EMBI).

q	CR	CAR	σ_{an}	SR	p-value vs.	
					GMV	$1/N$
0.00001	807.5%	31.7%	16.2%	0.8937	0.27	0.00
0.00005	804.9%	31.7%	16.2%	0.8926	0.28	0.00
0.0001	801.2%	31.6%	16.2%	0.8913	0.28	0.00
0.0005	761.5%	30.9%	15.9%	0.8863	0.29	0.00
0.001	711.2%	29.9%	15.5%	0.8815	0.29	0.00
0.005	515.6%	25.5%	14.0%	0.8263	0.28	0.00
0.01	471.9%	24.4%	14.3%	0.7743	0.06	0.00

This table reports the cumulative out-of-sample returns by the end of 2013 (CR), compound annual returns (CAR), annual standard deviation (σ_{an}), and the out-of-sample Sharpe ratio for the v=0.5n B–L portfolio for seven different values of q. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the Sharpe ratios of the GMV portfolio (0.8337) and the $1/N$ portfolio (0.3782), respectively (S&P EMBI).

Figure 1.8: Cumulative Out-of-Sample Returns of the B–L 0.5n Portfolio for Different q (MSCI).



This figure shows the cumulative out-of-sample quarterly returns for the v=0.5n B–L portfolio for seven different values of q from 2006 until 2013. The portfolios are readjusted quarterly (MSCI).

Table 1.11: Variations of q (MSCI).

q	CR	CAR	σ_{an}	SR	p-value vs.	
					GMV	$1/N$
0.00001	243.6%	14.7%	13.7%	0.4887	0.06	0.03
0.00005	240.3%	14.6%	13.7%	0.4862	0.06	0.03
0.0001	236.6%	14.4%	13.6%	0.4834	0.06	0.03
0.0005	200.2%	13.0%	13.1%	0.4496	0.10	0.05
0.001	169.6%	11.7%	12.8%	0.4084	0.21	0.08
0.005	151.7%	10.8%	13.5%	0.3608	0.09	0.14
0.01	149.8%	10.7%	13.8%	0.3511	0.09	0.16

This table reports the cumulative out-of-sample returns by the end of 2013 (CR), compound annual returns (CAR), annual standard deviation (σ_{an}), and the out-of-sample Sharpe ratio for the v=0.5n B–L portfolio for seven different values of q. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the Sharpe ratios of the GMV portfolio (0.3852) and the $1/N$ portfolio (0.2645), respectively (MSCI).

q. As shown in Table 1.11 the same effect applies to the $CARs$. As for the S&P EMBI data set the effect of the values for the elements of q on σ_{an} is ambiguous.

The effect of the value of the elements of q on the Sharpe ratio of the B–L portfolio is also strong. The Sharpe ratio increases with a reduction of the values of the elements of q. For q values below 0.0005 the Sharpe ratios are between 0.4496 and 0.4887 and are statistically significant larger at the 0.10 level than the Sharpe ratio of the GMV portfolio. For values of 0.0005 and 0.001 for the elements of q the Sharpe ratio of the B–L portfolio is still larger than that of the GMV portfolio, however the difference is statistically insignificant. For the two largest values for the elements of q the Sharpe ratio of the B–L portfolio is smaller than that of the GMV portfolio. For all chosen values for the elements of q the Sharpe ratio of the B–L portfolio is larger than that of the $1/N$ portfolio. However, the difference is statistically insignificant for the two largest chosen values for the elements of q.

The explanation for the strong effect that the values of the element of q have on the CRs and Sharpe ratios is quite intuitive. We use the view vector q to reduce the holdings in so-called dead assets. The holdings in an asset that is involved in the views (identified with the matrix P) decrease when the elements of q become smaller. Since we do not allow negative holdings, the smallest holding in an asset is zero. Therefore the effect of a further reduction of the elements of q becomes very small at a certain level, because the holdings in most assets involved in a view have already reached zero. If the values for the elements of q get quite large (such as 0.01), we induce the opposite effect of our initial goal. The holdings in the assets with a view, the dead assets, increase and therefore the CR of the portfolio decreases.

1.4.3.3 Variations of the Time Period

In Section 1.4.1 and Section 1.4.2 we showed the superior performance of the B–L portfolio with $v = 0.5n$ compared to the GMV portfolio and the $1/N$ portfolio in terms of CRs by the end of 2013 and of Sharpe ratios. However, the good performance of the B–L portfolio might be induced by the specific time horizon. Therefore, we assess whether the superior results of the B–L portfolio hold up for different time periods. In particular it is interesting to check the performance differences between the B–L portfolio with $v = 0.5n$, the GMV portfolio, and the $1/N$ portfolio for periods with low, normal, and high returns.

We assessed the performance of our model for several five-year periods with different starting dates. We chose a five-year time horizon to provide periods with a meaningful length on the one hand and a reasonable number of different periods on the other. For the S&P EMBI data set we used 13 different starting dates from January 2006 until January 2009. For the MSCI data set we used 17 different starting dates from January 2005 until January 2009. Besides this we implemented our model as described in Section 1.3, which includes a readjustment of the portfolio every 90 days.

In Table 1.12 we report CRs and out-of-sample Sharpe ratios for the B–L portfolio with $v = 0.5n$, the GMV portfolio, and the $1/N$ portfolio for the S&P EMBI data set for 13 different time periods. We also provide p-values of the difference between the Sharpe ratio of the B–L portfolio and those

Table 1.12: Variations of Time Periods (S&P EMBI).

Time	*CR BL*	*CR* GMV	*CR* $1/N$	*SR BL*	*SR* GMV	*SR* $1/N$	p-value vs.	
							GMV	$1/N$
1–20	364.2%	282.4%	231.0%	0.8891	0.8157	0.4208	0.27	0.01
2–21	285.3%	223.6%	192.9%	0.7807	0.7168	0.3819	0.29	0.02
3–22	323.8%	249.9%	212.9%	0.8699	0.7959	0.4104	0.27	0.01
4–23	296.4%	216.8%	168.5%	0.8473	0.7390	0.3603	0.20	0.01
5–24	216.3%	170.5%	114.7%	0.7458	0.6653	0.2903	0.27	0.01
6–25	204.4%	167.7%	128.4%	0.7391	0.6658	0.3128	0.28	0.01
7–26	183.9%	140.9%	75.5%	0.7221	0.6260	0.2325	0.24	0.01
8–27	191.2%	149.5%	70.6%	0.7271	0.6483	0.2269	0.27	0.01
9–28	199.0%	155.5%	78.1%	0.7542	0.6777	0.2444	0.28	0.01
10–29	212.3%	175.9%	110.7%	0.7940	0.7553	0.3031	0.39	0.01
11–30	217.9%	179.0%	109.9%	0.8085	0.7678	0.3035	0.38	0.01
12–31	197.3%	184.4%	147.4%	0.7301	0.8019	0.3724	0.29	0.05
13–32	289.2%	248.4%	237.0%	1.1641	1.1548	0.5163	0.48	0.01

This table reports the cumulative out-of-sample returns (CR) and out-of-sample Sharpe Ratios (SR) for the B–L portfolio with $v = 0.5$, the GMV portfolio, and the $1/N$ portfolio for 13 different time periods of five years each. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the GMV portfolio or the $1/N$ portfolio, respectively (S&P EMBI).

of the two other portfolios for each period. The numbers under the heading Time indicate the time period, with 1 being the first adjustment date in January 2006 and 32 the final adjustment date in October 2013. We can see that the B–L portfolio has a larger CR than the GMV portfolio and the $1/N$ portfolio in each of the 13 time periods. For example, in the period with the lowest returns—between July 2007 and June 2011 (indicated by 7–26 in Table 1.12)—the B–L portfolio has a CR of 183.9% compared to CRs of 140.9% for the GMV portfolio and 75.5% for the $1/N$ portfolio. The same applies for the period with the highest returns—between January 2006 and December 2010 (indicated by 1–20 in Table 1.12). In this period the B–L portfolio has a CR of 364.2% compared to CRs of 282.4% for the GMV portfolio and 231.0% for the $1/N$ portfolio. The B–L portfolio also has a larger Sharpe ratio than the $1/N$ portfolio in each of the 13 periods. The difference is statistically significant at the 0.05 level in each period. The B–L portfolio has a larger Sharpe ratio than the GMV portfolio for 12 out of the 13 periods, however the difference is statistically insignificant for each of these periods.

In Table 1.13 we report CRs and out-of-sample Sharpe ratios for the B–L portfolio with $v = 0.5n$, the GMV portfolio, and the $1/N$ portfolio for the MSCI data set for 17 different time periods. We

Table 1.13: Variations of Time Periods (MSCI).

Time	*CR BL*	*CR* GMV	*CR* 1/*N*	*SR BL*	*SR* GMV	*SR* 1/*N*	*p*-value vs. GMV	*p*-value vs. 1/*N*
1–20	101.3%	56.6%	52.5%	0.3770	0.2382	0.1665	0.00	0.05
2–21	96.4%	56.7%	54.8%	0.3662	0.2414	0.1739	0.01	0.06
3–22	80.1%	47.1%	46.4%	0.3210	0.2058	0.1557	0.01	0.08
4–23	76.9%	46.3%	33.5%	0.3155	0.2070	0.1247	0.02	0.06
5–24	78.5%	46.0%	33.5%	0.3255	0.2114	0.1275	0.02	0.05
6–25	60.9%	38.9%	37.8%	0.2729	0.1870	0.1424	0.07	0.15
7–26	72.6%	46.6%	44.8%	0.3270	0.2337	0.1655	0.06	0.11
8–27	53.9%	32.1%	19.4%	0.2615	0.1722	0.1003	0.07	0.10
9–28	37.5%	18.7%	5.4%	0.2051	0.1102	0.0597	0.07	0.11
10–29	33.2%	20.3%	18.9%	0.1927	0.1277	0.1075	0.16	0.24
11–30	35.6%	21.6%	2.6%	0.2136	0.1441	0.0614	0.15	0.11
12–31	48.3%	29.1%	18.3%	0.2839	0.1956	0.1140	0.10	0.09
13–32	44.4%	25.9%	17.9%	0.2704	0.1834	0.1158	0.10	0.11
14–33	76.3%	55.5%	54.5%	0.4233	0.3581	0.2113	0.18	0.05
15–34	73.1%	66.6%	64.5%	0.4162	0.3961	0.2335	0.40	0.08
16–35	74.7%	61.8%	77.7%	0.4262	0.3752	0.2637	0.28	0.10
17–36	100.8%	92.8%	159.4%	0.6018	0.5989	0.4357	0.49	0.15

This table reports the cumulative out-of-sample returns (CR) and out-of-sample Sharpe Ratios (SR) for the B–L portfolio with $v = 0.5$, the GMV portfolio, and the $1/N$ portfolio for 13 different time periods of five years each. The sub-columns under the heading p-value vs. show the p-values of the difference between the Sharpe ratio of the B–L portfolio and the GMV portfolio or the $1/N$ portfolio, respectively (MSCI).

can see that the B–L portfolio has a larger CR than the GMV portfolio in each of the 17 periods. It also has a larger CR than the $1/N$ portfolio in 15 out of the 17 periods while it is only smaller in the periods between October 2008 and September 2013 and between January 2009 and December 2013 (indicated by 16–35 and 17–36 in Table 1.13, respectively). The B–L portfolio has a larger Sharpe ratio than the $1/N$ portfolio in each of the 17 periods. The difference is statistically significant at the 0.10 level in 10 out of the 17 periods (for the other periods the p-values are between 0.11 and 0.24). The B–L portfolio also has a larger Sharpe ratio than the GMV portfolio in each of the 17 periods. The difference is statistically significant at the 0.10 level in 11 out of the 17 periods.

1.4.4 Portfolio Composition

In the Introduction we mentioned the appearances of extreme positions as an undesirable feature of mean-variance portfolios (see He and Litterman (1999)). In the last step of our results section we analyze if such extreme positions also occur in our modified B–L portfolios. Since we have a large number of candidate assets (591 for the S&P EMBI data set and 940 for the MSCI data set), portfolios with a relatively large number of assets, and many portfolio adjustments (32 for the S&P EMBI data set and 36 for the MSCI data set) we do not look at single composition in detail and rather analyze the big picture. We look at the average number of assets in the portfolio over all adjustment points and the average holding in the most represented asset, the five most represented assets, and the ten most presented assets in the portfolio. Furthermore we introduce an explicit portfolio diversification measurement.

We use the complement of the Herfindahl index (HI)—one of the most widely used measures of economic concentration—as our explicit diversification measurement (see Woerheide and Persson (1993)). This diversification index DI is proposed by Woerheide and Persson (1993), who consider its explanatory power adequate for general use:

$$DI = 1 - HI = 1 - \sum_{j=1}^{n} {w_{jf}^{*}}^{2}. \tag{1.25}$$

The index has the nice feature that a portfolio with no diversification, consisting only of one asset, would have a DI value of zero, while the perfectly diversified portfolio would have a value of 1. All other portfolios have a value somewhere in between.

Table 1.14 provides information about the composition of the eight B–L portfolios for the S&P EMBI data set. The average number of assets in the portfolio over all adjustment points decreases from 52 for the portfolio with $v = 0.25n$ to 25 for the portfolio with $v = 0.6n$. The average holding in the most represented asset increases from 16.5 percent for the portfolio with $v = 0.25n$ to 32.0 percent for the portfolio with $v = 0.55n$. An exception is the portfolio with $v = 0.6n$, which has a slightly smaller average holding (27.0%) in its most represented asset than the $v = 0.5n$ and the $v = 0.55n$ portfolios. A similar effect can be found for the five and ten most represented assets in the portfolio. Their share in the portfolio increases with an increase of v, from 40.8% (five most represented) and 60.3% (ten most represented) for the portfolio with $v = 0.25n$, to 66.9% and 86.8%, respectively, for the portfolio with $v = 0.6n$. The DI decreases from 94.2% for the portfolio with

Table 1.14: Portfolio Composition (S&P EMBI).

Portfolio	Assets	Top 1	Top 5	Top 10	DI
$v = 0.25n$	52	16.5%	40.8%	60.3%	94.2%
$v = 0.3n$	43	19.0%	44.9%	65.9%	93.0%
$v = 0.35n$	39	21.0%	48.8%	70.7%	91.9%
$v = 0.4n$	35	22.7%	52.4%	74.7%	90.9%
$v = 0.45n$	31	25.2%	57.5%	79.6%	89.3%
$v = 0.5n$	29	28.6%	62.5%	83.8%	87.3%
$v = 0.55n$	27	32.0%	66.8%	86.3%	85.2%
$v = 0.6n$	25	27.0%	66.9%	86.8%	86.4%

This table reports the average portfolio compositions for the eight B–L portfolios over all 32 adjustment points. The "Assets" column indicates the average number of assets in the portfolio. "Top 1" is the average holding in the most represented asset in the portfolio. "Top 5" is the average holdings in the five most represented assets in the portfolio. "Top 10" is the average holdings in the ten most represent assets in the portfolio. "DI" is the diversification index (S&P EMBI).

$v = 0.25n$ to 85.2% for the portfolio with $v = 0.55n$. Afterward it increases again slightly to 86.4% for the portfolio with $v = 0.6n$.

Table 1.15 provides information about the composition of the eight B–L portfolios for the MSCI data set. There are several similarities to the S&P EMBI data set. The number of assets and the DI decrease with an increase in v while the average holding in the most, the five most, and ten most represented assets increases. However, there are also some differences. The number of assets in the portfolio and the holdings in the most represented asset are considerably larger in the S&P EMBI data set for every value of v. The large holding in the most represented asset also leads to worse DIs for every v for the S&P EMBI data set.

In general the B–L portfolios for both data sets are not perfectly diversified; the large holding in one asset for the S&P EMBI data set, especially, is troubling. However, all B–L portfolios do at least have 20 assets in their composition and they do not have extreme positions as most mean-variance portfolios do. Furthermore 11 out of the 16 portfolios have a DI of at least 90 percent. Additionally an improvement in portfolio diversification could be reached easily by introducing maximum weight constraints when solving the maximization problem (1.11).

1.5 Conclusion

In this paper, we have developed a modified Black-Litterman model using the GMV portfolio as a reference portfolio. The model mostly relies on variance-covariance estimations as input parameters,

Table 1.15: Portfolio Composition (MSCI).

Portfolio	Assets	Top 1	Top 5	Top 10	DI
$v = 0.25n$	34	9.4%	40.0%	64.4%	94.6%
$v = 0.3n$	31	10.2%	43.3%	68.5%	94.1%
$v = 0.35n$	28	11.2%	47.2%	74.1%	93.3%
$v = 0.4n$	25	12.4%	51.1%	78.8%	92.4%
$v = 0.45n$	24	13.4%	54.4%	82.3%	91.7%
$v = 0.5n$	22	14.7%	57.5%	85.2%	91.0%
$v = 0.55n$	21	16.0%	60.2%	87.2%	90.4%
$v = 0.6n$	21	17.7%	63.5%	88.6%	89.5%

This table reports the average portfolio compositions for the eight B–L portfolios over all 32 adjustment points. The "Assets" column indicates the average number of assets in the portfolio. "Top 1" is the average holding in the most represented asset in the portfolio. "Top 5" is the average holdings in the five most represented assets in the portfolio. "Top 10" is the average holdings in the ten most represent assets in the portfolio. "DI" is the diversification index (MSCI).

estimations of the exact values of expected returns are not necessary. The introduction of a general rule, based on relative expected returns between the assets and the systematic risk β of the assets, to simulate investors' views enables us to remove dead assets from the GMV portfolio. In the context of our general rule we were also able to remove from the original B–L setup two parameters whose values are the subject of significant controversy in the literature. Overall we were able to combine two of the most popular portfolio optimization models and reap the benefits of both.

We implemented the model using two different real-life data sets and demonstrated that our approach works in a large-scale setup. We benchmarked the cumulative returns of the portfolios calculated with our model against the GMV portfolio and the $1/N$ portfolio and showed that our modified B–L portfolios outperform these two other portfolios considerably in terms of CR and out-of-sample Sharpe ratios. A detailed sensitivity analysis indicated the importance of the risk aversion parameter δ and the elements of the view vector q. A larger value for the risk aversion and smaller values for the elements of q led to clearly larger cumulative returns. An analysis for different time periods showed that our modified B–L model has—aside from one single period in one of the data sets—a higher out-of-sample Sharpe ratio than the GMV portfolio. The difference in Sharpe ratios is statistically significant in almost half of the periods. Additionally our model has a higher Sharpe ratio than the $1/N$ portfolio in each period for both data sets with a statistically significant difference in almost all periods.

This paper presents the first extensive analysis of a B–L model with the GMV portfolio as the reference portfolio. We strongly believe that the combination of these two famous portfolio optimization models could be beneficial for asset management companies in practice. On the one hand, the GMV portfolio—as a starting point—is clear, intuitive, and understandable for potential customers. On the other, our modified B–L approach removes dead assets from the GMV portfolio and provides high and reliable returns at a reasonable risk exposure, which is expressed in the strong out-of-sample Sharpe ratios. Furthermore, our general rule and the elimination of the two most uncertain parameters simplifies the implementation of the B–L approach considerably. In addition we were able to show that our model is also implementable with a large set of candidate assets, which makes it suitable for practical use.

Essay 2

Percentile Forecasting

Percentile Forecasting and Trading Strategies[1]

Maximilian Adelmann
Dept. of Business Administration
University of Zurich
Moussonstrasse 15
8044 Zurich, Switzerland
maximilian.adelmann@business.uzh.ch

Cyril Bachelard
OLZ & Partners
Asset and Liability Management AG
Marktgasse 24
3011 Bern, Switzerland
cyril.bachelard@olz.ch

November 2016

Abstract

Recent additions to the forecasting literature have shown that the sign of returns can—in contrast to the mean of returns—be predicted. This paper presents a binary logit sign prediction approach. As a novelty in the literature we extend this approach to forecast also the probabilities for returns larger than several percentiles. Our binary logit percentile approach includes four different models that differ in terms of their explanatory variables. All four models comprise the conditional volatility as an explanatory variable. The second model adds unconditional skewness, the third adds the turbulence contribution based on the Mahalanobis distance, and the fourth adds both unconditional skewness and the turbulence contribution as explanatory variables. We present several common tests to assess the predictive power of our models and also develop a new test to visualize our test results. A numerical application of our binary logit approach to empirical data sets demonstrates the superior forecasting performance of Model 3 and Model 4 and return percentiles smaller than zero. Based on our percentile forecasts we develop trading strategies that provide Sharpe ratios superior to those of their respective benchmark strategies.

Keywords: GARCH, out-of-sample tests, percentile forecasting, sign forecasting, trading strategies.

[1]We thank Lorenz Beyeler and Carmine Orlacchio for helpful discussions on the subject and comments on earlier implementations of the model. We are grateful to seminar audiences at OLZ & Partners for helpful comments. We would also like to thank János Mayer and Karl Schmedders for their support and guidance on this project.

2.1 Introduction

One main challenge for asset managers—besides the choice of a successful portfolio optimization approach—is the correct estimation of expected stock market returns. Merton (1980) and Chopra and Ziemba (1993) show that such mean estimations are far less accurate than variance-covariance estimations. Christoffersen and Diebold (2006) even state that approximate conditional mean independence and hence little or no mean "forecastability" is a prominent phenomenon in the finance literature. However, they explain that asset return signs are predictable even if we are not able to forecast means. Several studies (see for example Leung et al. (2000), Nyberg (2011), Chevapatrakul (2013), and Nyberg and Pönkä (2015)) provide empirical evidence of signs being forecastable to a certain extent.

The idea behind sign forecasting is to provide a probability—expressed as a percentage—of a return larger than 0 percent in the next period. Thereby mean forecasts (which are not or are only poorly predictable) are not needed to determine the probability of a positive return. Instead, other variables—which are predictable—are forecasted. These explanatory variables can include, for example, volatility (see Christoffersen and Diebold (2006)), skewness and kurtosis (see Christoffersen et al. (2007)), the interest rate and the term spread (see Chevapatrakul (2013)), or a recession indicator (see Nyberg (2011)). The probability of a positive return is determined in a forecasting model based on the explanatory variables. Suitable forecasting models include, for example, classification models such as logit and probit models.

Sign forecasts can be used to develop investment strategies. A common application in the literature—see Pesaran and Timmermann (1992), Leung et al. (2000), Nyberg (2011), and Chevapatrakul (2013)—is trading strategies between stock market investments and a safer option such as bonds or treasury bills. In the simplest form of such a strategy, investments are made in the stock market if the probability of a positive return of the stock market is larger than 50 percent and investments are made in the safer asset if the probability is smaller than 50 percent.

In this paper, we present a binary logit model for forecasting the sign of asset returns based on the approach of Chevapatrakul (2013). In a next step we extend this sign forecasting model to a percentile forecasting model. In particular we predict the probabilities for returns larger than -2%, -4%, -6%, and -8% additionally to the sign forecasting. We take given return levels (for example $R_1 = 0\%$, $R_2 = -2\%$, $R_3 = -4\%$, etc.) and compute the corresponding probabilities for these return levels. This is a reverse reasoning to statistics where we have a probability level given and determine the expected return for this level. We are the first to extend sign forecasting to percentile forecasting in this fashion. Predictions that provide probabilities for returns larger than certain percentiles can be useful for an investor who particularly wants to avoid periods with large negative returns and is willing to miss out on positive returns in other periods in exchange. Furthermore, we assume that negative percentiles should be easier to forecast than signs. According to Christoffersen and Diebold (2006) we need non-zero means to forecast signs based on volatility forecasts. This implies that we would, for example, need a mean that is not equal to -2% if we forecast the probability for a return larger than -2 based on volatility. Since a mean equal to -2% is in general less likely than a zero mean, a forecast for the -2% percentile should on average be possible more often than a sign forecast.

After the formulation of our percentile forecasting approach we explain and justify our choice of explanatory variables. For our first model we only use the conditional volatility as an explanatory variable, in line with Christoffersen and Diebold (2006). For our second model we additionally use skewness, as do Christoffersen et al. (2007). For our third model we consider, in addition to the conditional volatility, an explanatory variable that has not been used in the sign forecasting literature before: the turbulence contribution based on the Mahalanobis distance. The Mahalanobis distance was originally introduced to analyze the human skull, but can also be used as a suitable measurement for financial turbulence (see Kritzman and Yuanzhen (2010)). Intuitively such a turbulence measurement is also promising for forecasting percentiles. In our fourth model we use the conditional volatility, the skewness, and the turbulence contribution as explanatory variables.

After the theoretical formulation of our percentile forecasting model and the choice of explanatory variables, we numerically implement the models. We use three real-life data sets, which we received from OLZ & Partners Asset and Liability Management AG (OLZ) in Bern/Switzerland. The first two data sets include assets from developed markets and the third from emerging markets. We execute the sign forecast for every asset individually for 51 different time periods. The conditional volatilities and the unconditional skewness are calculated using a GARCH(1,1) approach.

In the next step we present several tests for assessing the quality of our percentile forecasts. In particular we evaluate the performance of our model using two tests that are common in the sign prediction literature. The first of these is a pseudo-R^2 measure, a method that is for example also used by Chevapatrakul (2013) and Nyberg and Pönkä (2015). The second is the Brier(Abs) score introduced by Christoffersen et al. (2007). Additionally we present our own forecast evaluation test that is well suited to visualizing the predictions.

In a last step we introduce a trading strategy based on our percentile forecasts. In this strategy we invest in the stock market or the risk-free rate based on the percentile forecast for all assets. We analyze the investment performance for all models, different percentiles, and all data sets.

The Literature

Research on asset return sign predictability is still rather scant. Leung et al. (2000) evaluate the ability of several classification and level estimation models to predict the sign of stock indices. Their empirical investigations suggest that classification models outperform level estimation models in terms of forecasting the sign of the stock market. The classification techniques tested for sign forecasting by Leung et al. (2000) are linear discriminant analysis, probabilistic neural network , logit, and probit. The last two of these models are also applied in other research studies. Chevapatrakul (2013) develops a research method based on two stages. In the first stage he estimates explanatory variables based on a bivariate BEKK-GARCH approach, which he uses in a binary logit model to predict the sign of the UK stock market return in the second stage. Nyberg (2011) and Nyberg and Pönkä (2015) both apply probit models for sign predictions. Nyberg (2011) develops a binary dependent dynamic probit model to forecast the sign of the monthly returns of the US stock market. He includes the recession forecast as an explanatory variable in his forecasting method, which enables him to outperform other predictive models. Nyberg and Pönkä (2015) use univariate and bivariate probit models to predict the signs of international stock markets with a focus on the United States. Their in-sample and out-of-sample forecasting results provide evidence of the superior predictive power of a new bivariate probit model that allows for a predictive linkage between markets.

In one of the most influential papers in the sign prediction literature, Christoffersen and Diebold (2006) show a direct and intuitive link between volatility and sign forecastability. They demonstrate that sign forecasts can be made based on volatility forecasts—as long as expected returns are non-zero—even in the presence of mean independence (and hence no mean forecastability). Based on these findings, Christoffersen et al. (2007) develop two direction-of-change prediction models. The first one makes direct use of the analysis of Christoffersen and Diebold (2006) and includes solely the conditional volatility as an explanatory variable. The second one uses, additionally, the skewness and kurtosis. Christoffersen et al. (2007) evaluate the performance of their two models using empirical evidence.

Suitable predictive-power evaluation tests are essential to assess the performance of forecasting models. One popular method of evaluating predictive power is market timing tests. These tests (see for example Henriksson and Merton (1981)) are in general concerned with signs and sign forecasting. Pesaran and Timmermann (1992) specifically propose a test to predict directional accuracy, while Pesaran and Timmermann (2009) suggest another predictability test as an extension to the former test statistic. Anatolyev and Gerko (2005) introduce a new excess predictability test based on the market timing method. Another set of tests for assessing predictive power are Brier tests. Diebold and Lopez (1996) suggest the traditional $Brier(Sq)$ measurement for evaluating forecast performance, which Christoffersen et al. (2007) slightly modify to create the Brier(Abs) measurement. These Brier scores are widely used to assess predictive power in empirical studies (see for example Christoffersen et al. (2007), Anatolyev (2009), and Nyberg and Pönkä (2015)). A common set of tests for assessing the performance of logit and probit models are pseudo-R^2 measurements. Such Pseudo-R^2 measures are for example used in empirical studies by Chevapatrakul (2013) and Nyberg and Pönkä (2015).

Another way to assess the out-of-sample performance of sign forecasts—and one of the main reasons to execute them in the first place—is trading strategies based on direction-of-change predictions. Several simple trading strategies—in which investments are made in either the stock market or in a safer alternative such as bonds or treasury bills—are implemented and assessed empirically by Pesaran and Timmermann (1995), Leung et al. (2000), Nyberg (2011), Chevapatrakul (2013), and Nyberg and Pönkä (2015).

The remainder of this paper is organized as follows. We introduce our binary logit percentile forecasting model and justify our choice of explanatory variables in Section 2.2. In Section 2.3 we describe the three data sets we use and explain the calculation of our explanatory variables.

Section 2.4 presents our forecast evaluation tests and the specific results of our data sets for these tests. In Section 2.5 we show a trading strategy approach that is based on our percentile forecasts and analyze the performance of this strategy. Section 2.6 concludes.

2.2 Forecasting Model

In this section we present a sign forecasting model and extend it to our percentile forecasting model. In the first subsection we explain the general setting of our binary logit model. In the second subsection we show the specification of this model and justify our choice of explanatory variables.

2.2.1 Binary Logit Model

Table 2.1: Percentile Forecasting Parameter.

Symbol	Description
$k = 1, \ldots, K$	Forecasting periods
k	Present date, when forecast is made
p	Percentile
v	Number of explanatory variables
$Pr(R_{k+1} \geq 0 \mid X_k)$	Conditional probability for a positive return at $k+1$
$Pr(R_{k+1} < 0 \mid X_k)$	Conditional probability for a negative return at $k+1$
$Pr(R_{k+1} \geq p \mid X_k)$	Conditional probability for a return larger p at $k+1$
$Pr(R_{k+1} < p \mid X_k)$	Conditional probability for a return smaller p at $k+1$
X_k	$(v+1) \times 1$ vector of explanatory variables at time k
$\hat{\beta}_S$	$(v+1) \times 1$ vector of estimated parameters (sign forecasting)
$\hat{\beta}_P$	$(v+1) \times 1$ vector of estimated parameters (percentile forecasting)
σ_k	Conditional standard deviation at time k
γ_k	Unconditional skewness at time k
D_k	Turbulence contribution at time k

This table provides a parameter overview for the percentile forecast models. The respective predicted probabilities are denoted $\widehat{Pr}$.

To the best of our knowledge no percentile forecasting model that takes given return levels and computes the corresponding probabilities can be found in the literature.[2] Hence, we first provide an overview of existing sign forecasting models and extend one of these approaches to our percentile

[2] The term "percentile forecasting" has been sparsely used before, but in all cases with a completely different meaning from the terminology used in this paper.

forecasting model. Leung et al. (2000) present and compare the forecasting performance of several different classification models and level estimation models and show the superiority of the classification models in terms of predicting the direction of the stock market. The classification approaches analyzed by Leung et al. (2000) are linear discriminant analysis, logit, probit, and probabilistic neural network. All these models have in common that they predict direction of change based on probability. Nyberg (2011) and Nyberg and Pönkä (2015) use binary dependent dynamic and bivariate probit models for sign predictions, while Chevapatrakul (2013) uses a binary logit model to forecast return signs. Binary probit and logit models have the advantageous feature that their setup can easily be extended from sign forecasting to percentile forecasting. This is not the case for the direction-of-change forecasting models used by Christoffersen and Diebold (2006) and Christoffersen et al. (2007). Therefore we tested both a binary logit and a binary probit forecasting approach. The numerical results were very similar, with a slightly better performance from the logit model.

Based on this superior numerical performance we chose a binary logit model for our prediction approach and formulate a sign forecasting approach similar to the models used by Chevapatrakul (2013). We have a vector X_k (with the first element equal to 1) of explanatory variables that are available at time k. We define the probability of a positive return in period $k+1$ conditional on X_k by the logit cumulative density function

$$Pr(R_{k+1} \geq 0|X_k) = \frac{e^{X_k'\beta_S}}{1 + e^{X_k'\beta_S}} \tag{2.1}$$

and $Pr(R_{k+1} < 0|X_k) = 1 - Pr(R_{k+1} \geq 0|X_k)$, where β_S are the unknown parameters to be estimated by maximum likelihood. We can calculate the predicted probabilities $\widehat{Pr}(R_{k+1} \geq 0|X_k)$ and $\widehat{Pr}(R_{k+1} < 0|X_k)$ once we obtain the estimated parameters $\hat{\beta}_S$.

As we explained above, the forecasting literature focuses on direction-of-change forecasts (see for example Christoffersen and Diebold (2006), Christoffersen et al. (2007), Chevapatrakul (2013), Nyberg (2011), and Nyberg and Pönkä (2015)). For our model we also consider such a sign forecast. However, a major novelty of our approach is an additional forecast of the probabilities for returns larger than several percentiles. In particular we predict the probabilities for returns larger than $p = 0\%$ (which equals the sign forecast), $p = -2\%$, $p = -4\%$, $p = -6\%$, and $p = -8\%$. These percentile forecasts have the advantage—compared to the sign forecast—that they can be a good indicator for periods of significant negative returns and thereby be useful for investors who want to avoid investments in the stock market in such periods. And—following the argumentation of Christoffersen and Diebold (2006)—negative percentile forecasts are easier to conduct than the sign forecast because the negative percentile is more likely to differ from the expected return than is the sign forecast percentile $p = 0\%$. We focus on the prediction of negative percentiles because we are particularly interested in detecting assets and periods with a high probability of negative returns in order to reduce the portfolio risk. However, our model could be extended to also predict the probabilities of positive percentiles.

The formulation of our percentile forecasting model is very similar to the sign forecasting approach shown in (2.1). We define the probability of a return larger than p in period $k+1$ conditional on X_k

by the logit cumulative distribution function

$$Pr(R_{k+1} \geq p|X_k) = \frac{e^{X_k' \beta_P}}{1 + e^{X_k' \beta_P}} \tag{2.2}$$

and $Pr(R_{k+1} < p|X_k) = 1 - Pr(R_{k+1} \geq p|X_k)$, where β_P are the unknown parameters to be estimated by maximum likelihood.

2.2.2 Explanatory Variables for Our Logit Model

The choice of well-suited explanatory variables is essential for a good forecasting performance of our binary logit model shown in Section 2.2.1. We execute our percentile forecast for each asset separately (instead of for a whole index at once) and use explanatory variables that have different values for each asset. Christoffersen and Diebold (2006) show evidence that volatility dependence produces sign dependence as long as expected returns are non-zero. Hence, volatility, or more specifically standard deviation, appears to be a suitable explanatory variable for our percentile forecasting approach. We obtain the volatility forecasts using a GARCH approach, as do Christoffersen and Diebold (2006) and Chevapatrakul (2013). In particular we use the conditional standard deviation forecasted with GARCH(1,1) (we explain the GARCH approach in detail in Section 2.3.3) as an explanatory variable for our Model 1, denoted $X_{1,k}$:

$$X_{1,k} = (1, \sigma_k)'. \tag{2.3}$$

Christoffersen et al. (2007) extend the framework of Christoffersen and Diebold (2006) to forecasts that are also based on conditional skewness and kurtosis. We evaluated the additional influence of skewness and kurtosis on sign forecasting performance in several numerical tests and found that only skewness forecasts had a potentially positive effect on it. We show the detailed calculation of the skewness within the GARCH approach in Section 2.3.3. Since the calculation of the conditional skewness did not lead to beneficial results we use the unconditional skewness in addition to the conditional standard deviation as an explanatory variable for our Model 2, denoted $X_{2,k}$:

$$X_{2,k} = (1, \sigma_k, \gamma_k)'. \tag{2.4}$$

For our third model we introduce the turbulence contribution based on the Mahalanobis distance as an explanatory variable, which has to the best of our knowledge not been used in the sign forecasting literature so far. The Mahalanobis distance was originally developed by Mahalanobis (1927) to analyze human skulls. Kritzman and Yuanzhen (2010) show that it can also serve as a measurement of financial turbulence. Intuitively such a measurement (we describe the Mahalanobis distance in detail in Section 2.3.3) for detecting market turbulence appears promising for percentile forecasting and therefore we use the turbulence contribution in addition to the conditional standard deviation as an explanatory variable for our Model 3, denoted $X_{3,k}$:

$$X_{3,k} = (1, \sigma_k, D_k)'. \tag{2.5}$$

In our Model 4, denoted $X_{4,k}$, we use both the unconditional skewness and the turbulence contribution in addition to the conditional standard deviation as explanatory variables:

$$X_{4,k} = (1, \sigma_k, \gamma_k, D_k)'. \tag{2.6}$$

2.3 Data Sets and Implementation

In the first two parts of this section we show the data sets and the approach to the implementation of our models developed in Section 2.2. We use three different data sets for the implementation, two from developed and one from emerging markets.[3] In the third part of this section we clarify the calculation of the conditional standard deviation, the unconditional skewness, and the turbulence contribution based on the Mahalanobis distance, which serve as explanatory variables in our logit models.

2.3.1 Data Sets

Table 2.2: Data Sets.

	MSCI DM	MSCI USA	MSCI EM
Number of available assets n	807	192	231
Data frequency	Weekly		
First available data	February 1994		
Periods between first data and first forecast	$b = 1, \ldots, B$		
First forecast date k=1	February 2004		
Last available data	December 2015		
Forecasting frequency	Every 12 weeks		
Forecasting periods	$k = 1, \ldots, K$ with $K = 51$		

This table reports details for our three data sets. We use discrete weekly total returns with price information.

We implement our percentile forecast models using three data sets, as reported in Table 2.2. For all three sets we have data available from February 1994 until December 2015. The first 10 years of the data set between February 1994 and February 2004 ($b = 1, \ldots, B$) are solely used to collect information for the forecasting periods, starting with $k = 1$ in February 2004.

In order to run our econometric models we require consistent data sets without gaps. Therefore, we first clean our data, starting with adjustments for past corporate actions and other corporate

[3]We thank OLZ for providing the data sets. We are especially grateful to Lorenz Beyeler and Carmine Orlacchio for professional advice and helpful discussions regarding the topic.

events. We then exclude all assets that do not have price information at every point in time of the investigated period. Additionally we disregard stocks that have not been traded for more than two weeks in a row. Reference index membership over the full sample period is not a requirement. This cleaning process is necessary to ensure that the resulting universe comprises clean time series. For this cleanliness we pay the price of a survivorship bias. However, this poses no problem in our context since all the inference is done within the same selection and no comparison to a benchmark index is made.

We use the cleaned data sets to compute discrete weekly returns with price information (total return, i.e., accounting for dividend payments) based on Wednesdays' closing quotes. We execute this process for the three data sets reported in Table 2.2 and end up with a selection of 807 assets for the MSCI World universe (MSCI DM), 231 assets for the MSCI Emerging Markets (MSCI EM) universe, and 192 assets for the MSCI USA.[4]

2.3.2 Implementation

The implementation of our model consists of a three-step process that is applied initially and then repeated at each rebalancing date. The three steps are: (GARCH) parameter estimation, volatility prediction, and percentile computation. Our back-testing simulation starts in February 2004 at $k = 1$ using the last 10 years of available data ($b = 1, \ldots, B$). We make forecasts for a time horizon of 12 weeks (in the remainder of the paper we refer to this as *quarterly*) because sign dependence (and also percentile dependence) is more likely to appear at such an intermediate return horizon than for high-frequency of low-frequency returns (see Christoffersen and Diebold (2006)). In total we have $K = 51$ forecasting periods, the last one in August 2015. At each period k the computations are based on all the available data from February 1994 on—instead of a moving time window—to use as much available information as possible for our forecasting model (at period k the available information consists of the periods $1, \ldots, B$ and $1 \ldots, k$).

2.3.3 Calculation of Our Explanatory Variables

In this section we present additional information about the explanatory variables we presented in Section 2.2.2. In Section 2.3.3.1 we explain the concepts of volatility dynamics and depict the calculation of the conditional volatility and the unconditional skewness. In Section 2.3.3.2 we present the calculation of the Mahalanobis distance and the contribution of turbulence for each asset.

2.3.3.1 Conditional Volatility and Skewness

Before we show the calculation of the conditional volatility we first provide some background information about volatility dynamics. Let r_t be a time series of stock returns. Many models for financial

[4] The number of stocks in the three indices changes over time: by the end of December 2015 the MSCI DM included 1,653 assets, the MSCI EM 838 assets, and the MSCI USA 633 assets.

time series used in practice assume a data generating process of the form

$$r_t = \mu_t + \epsilon_t,$$

where ϵ_t takes the multiplicative form $\epsilon_t = \sigma_t z_t$ with $z_t \sim iid\ N(0,1)$ and σ_t is defined by some autoregressive process.

Typically, the mean process of weekly asset returns lacks significant autocorrelation and we set $\mu_t = \mu$. Jondeau et al. (2001) differentiate between two families of models to describe volatility dynamics: generalized autoregressive conditional heteroscedasticity (GARCH) and stochastic volatility. In the GARCH model family, volatility is described as an exact function of a given set of parameters

$$\epsilon_t = \sigma_t z_t \tag{2.7}$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{2.8}$$

where p denotes the order of the GARCH terms σ^2 and q is the order of the ARCH terms ϵ^2. In the stochastic volatility model family, volatility is described as a stochastic function

$$\epsilon_t = \sigma_t z_t = z_t e^{h_t/2} \tag{2.9}$$

$$h_t = \omega + \beta h_{t-1} + \nu_{t-1}, \tag{2.10}$$

where ν_t is an $iid\ N(0, \sigma_\nu^2)$ process. z_t and ν_t are usually assumed to be independent.[5]

The main difference between the two model families is that there is only one source of uncertainty coming from z_t in the GARCH framework whereas in the stochastic volatility model there is an additional source of randomness. GARCH does not allow for errors in the dynamics of volatility. The impact of a given return on the volatility is determined by the estimated parameters and is the same for all times.

Both GARCH and stochastic volatility come in many forms (Bollerslev (2008) identifies over 150 different GARCH models) accounting for higher order autocorrelations, asymmetries, and departure from normality. Although not considered here, it is worth noting that alternative models exist and could be applied in this context as well. In particular, so-called realized volatility models—which approximate volatility by the sum of squared returns obtained from a higher frequency—are popular. Daily volatility levels for example may be extracted from the sum of squared intraday returns.

We use a simple GARCH(1, 1) structure as given by (2.7) to determine our conditional volatility explanatory variable.[6] However, as noted by Jondeau et al. (2001), the residuals of a GARCH model are in general not distributed according to a Gaussian distribution, suggesting that additional phenomena take place that are not captured by the simple GARCH model.

[5]Dependence between the two creates volatility asymmetry, which may sometimes be a desirable effect.

[6]We have also experimented with asymmetric GARCH structures, which are described in Appendix 2.B. However, this had no influence on our results and we do not provide any output for them.

Hence, instead of the Normal distribution we use a mixture of a Normal and an Inverse Gaussian distribution: the Normal Inverse Gaussian (NIG) distribution. The distribution is very flexible and incorporates skewness and excess kurtosis. Therefore, it provides a potentially better fit to financial data. We refer to Eberlein and Prause (2002) for a thorough examination of the distributional properties and its utility in finance and provide a basic description in Appendix 2.A. This leads to our following model:

$$r_t = \mu + \epsilon_t \tag{2.11}$$

$$\epsilon_t = \sigma_t z_t \tag{2.12}$$

$$z_t \sim f_{NIG}(\mu, \delta, \beta, \alpha) \tag{2.13}$$

$$\sigma_t^2 = \omega + \alpha_0 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{2.14}$$

where the parameters of the NIG distribution μ, δ, β, and α denote location, scale, skew, and shape respectively.

We do not model the dynamics of higher moments explicitly. However, since we re-estimate all parameters at each rebalancing the current setup provides us with a time-varying estimate for skewness and kurtosis.[7]

2.3.3.2 Mahalanobis Distance and Turbulence Contribution

The Mahalanobis metric measures the distance of an observation (i.e. a point in space represented by the position vector x) to the centre of mass of a distribution P characterized by vectors of means μ and covariance matrix Σ (we assume that the covariance matrix Σ is nonsingular). It has the following form:

$$D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (dx - d\mu)}. \tag{2.15}$$

The metric takes the correlation of variables in the data set into account. In case that the covariance matrix is just a multiple of the unit matrix (i.e. the data set is made of uncorrelated series having the same variances) the Mahanalobis distance coincides with the normalized Euclidean distance. Chow et al. (1999) determine the statistical unusualness of a cross section of returns on the basis of their historical multivariate distributions. Building on that idea, Kritzman and Yuanzhen (2010) introduced the definition of financial turbulence as the square of the Mahalanobis distance. Turbulence, in this sense, arises when returns deviate from their typical past behavior, be it through extreme price moves or through the convergence (decoupling) of previously uncorrelated (correlated) assets. Rewriting the definition of financial turbulence in Kritzman and Yuanzhen (2010) as

$$D_M^2(x) = (x-\mu)^T \Sigma^{-1} (x-\mu) = \sum_{i=1}^{n} (x-\mu)_i \left[d\Sigma^{-1}(x-\mu) \right]_i \tag{2.16}$$

[7]The connection between the distributional parameters and the standardized moments mean, variance, skewness, and kurtosis—that we can use for sign and percentile predictions—is given in Appendix 2.A.

we can define the total contribution of the i^{th} asset to the square of the Mahalanobis distance as

$$TCD_M^2(x) = (x-\mu)_i \left[\Sigma^{-1}(x-\mu)\right]_i = \frac{(x-\mu)_n}{2} \frac{\partial}{\partial x_n} D_M^2(x) \tag{2.17}$$

and we define $\frac{\partial}{\partial x_n} D_M^2(x)$ as the marginal contribution of the i^{th} asset to the squared Mahalanobis distance. This marginal contribution (the turbulence contribution based on the Mahalanobis distance) serves as an explanatory variable for our logit model, denoted D_k.

2.4 Forecast Evaluation

This forecast evaluation section consists of two subsections. In the first, we discuss several predominant forecast evaluation tests in the literature, introduce a new test, and determine the best suited tests for our approach. In the second, we report the test results for our three data sets and four models. In addition we show graphs to visualize the quality of our forecasts.

2.4.1 Forecast Evaluation Tests

In this section we discuss several tests that are used to evaluate sign forecasts in the literature and argue, which ones are suited best for our percentile forecasting approach. In addition we introduce a new test, which is well suited to visualizing our forecasting results. A common method of evaluating forecasts are market timing tests (see, for example, Henriksson and Merton (1981), the directional accuracy test by Pesaran and Timmermann (1992), or the excess predictability test by Anatolyev and Gerko (2005)). However, none of these market timing tests are well suited to evaluating our percentile forecasting approach. Christoffersen and Diebold (2006) explain, the market timing tests only differentiate between forecasts larger than (and equal to) 0.5 and smaller than 0.5. Thereby a forecast of 0.5001 is treated equally as a forecast of 0.9999 while it is treated fundamentally different than a forecast of 0.4999. This method is in particular unsuited for our percentile forecasting approach when $p < 0$. For these percentiles the forecasts are almost exclusively larger than 0.5, which makes a market timing test useless.

Another common method for evaluating forecasting results is pseudo-R^2 measures. They are counterparts of the coefficient of determination (R^2) and are well suited to measuring goodness-of-fit. Following Chevapatrakul (2013) we present the pseudo-R^2 measurements of McFadden (1974) and Estrella (1998) (the latter is also used by Nyberg and Pönkä (2015)), which are defined as

$$R^2_{McFadden} = 1 - \frac{\ln L_{UN}}{\ln L_{RS}} \tag{2.18}$$

$$R^2_{Estrella} = 1 - \left(\frac{\ln L_{UN}}{\ln L_{RS}}\right)^{(-2/K)\ln L_{RS}}, \tag{2.19}$$

where $\ln L_{UN}$ is the log-likelihood function of the unrestricted model, $\ln L_{RS}$ is the log-likelihood function of the restricted model, and K is the length of the time series. Both measures are bounded by 0 and 1 as are conventional R^2 statistics.

According to Nyberg and Pönkä (2015) the pseudo-R^2 measures need to be supported by other test statistics. They suggest the quadratic probability score, which is also known as the Brier(Sq) score and is used by Diebold and Lopez (1996) and Christoffersen et al. (2007). It has the following form:

$$Brier(Sq) = \frac{1}{K}\sum_{k=1}^{K} 2(\widehat{Pr}(R_{k+1} > p|X_k) - I(R_{k+1}))^2, \tag{2.20}$$

where $I(R_{k+1})$—which is based on the actual return in period $k+1$—is equal to 1 in the case of a return larger than p and 0 in the case of a return smaller than p. The Brier(Sq) values lie between 0 and 2, with 0 being the perfect and 2 being the worst possible score. Christoffersen et al. (2007) indicate that a few incorrect scores can dominate a majority of correct forecasts in the Brier(Sq) score. Hence, they suggest a modification of the test, which they call Brier(Abs). It has the following form:

$$Brier(Abs) = \frac{1}{K}\sum_{k=1}^{K} \left|\widehat{Pr}(R_{k+1} > p|X_k) - I(R_{k+1})\right|. \tag{2.21}$$

This test—which is also used by Anatolyev (2009)—has a perfect score of 0 (as does the Brier(Sq) test), and its worst score is 1.

All the tests that we have shown above provide a forecasting score for all forecasts at once. However, we are also interested in the performance of each estimated probability percentile (which should not be confused with percentile p). Therefore we introduce our own forecast estimation test that compares the estimated probabilities for returns larger than p with the average of the actual returns larger than p for each forecasted percentile separately. At first we put all forecasts for returns larger than p in buckets c depending on their percentile; for example all forecasts that predict a return larger than p with a probability of between 55 percent and $55.\overline{9}$ percent are put in a bucket, all forecasts between 56 percent and $56.\overline{9}$ percent are put in the next bucket, etc. Afterward we calculate for each bucket the average of the actual returns AAR_c that is larger than p with the following equation:

$$AAR_c = \frac{1}{n_c}\sum_{i=1}^{n_c} z_{i,k+1}, \tag{2.22}$$

where n_c is the number of assets in each bucket and $z_{i,t+1}$—which is based on the actual return in period $k+1$ for all assets i in this bucket c—is equal to 1 in the case of a return larger than p and 0 in the case of a return smaller than p. This test has the advantageous feature that it is well suited to a meaningful visual representation of the forecasting performance, with the bucket value for estimated returns larger than p on the x-axis and its respective AAR_c on the y-axis.

2.4.2 Test Results

In this section we present the test results for our four models. As explained in Section 2.4.1, market timing tests are not well suited for our percentile forecasts. Hence, we did not execute any market timing tests and focus on pseudo-R^2 measures, Brier scores, and our own test instead.

Table 2.3: McFadden pseudo-R^2.

		0%	−2%	−4%	−6%	−8%
DM	Model 1	0.0120	0.0164	0.0226	0.0302	0.0391
	Model 2	0.0219	0.0265	0.0334	0.0417	0.0518
	Model 3	0.0241	0.0297	0.0376	0.0473	0.0587
	Model 4	0.0334	0.0393	0.0480	0.0584	0.0709
USA	Model 1	0.0136	0.0197	0.0275	0.0368	0.0469
	Model 2	0.0239	0.0301	0.0385	0.0485	0.0592
	Model 3	0.0254	0.0337	0.0451	0.0587	0.0742
	Model 4	0.0358	0.0444	0.0568	0.0713	0.0876
EM	Model 1	0.0094	0.0124	0.0166	0.0211	0.0267
	Model 2	0.0246	0.0277	0.0322	0.0374	0.0437
	Model 3	0.0202	0.0244	0.0298	0.0358	0.0425
	Model 4	0.0345	0.0386	0.0442	0.0505	0.0578

This table reports McFadden pseudo-R^2 measures for five different values of p for the four models.

The scores of the pseudo-R^2 measures of McFadden (1974) and Estrella (1998) are rather similar for all models. Therefore we only report the scores for the pseudo-R^2 suggested by McFadden (1974) because it is the better known measure of the two (see Chevapatrakul (2013)). Since we have forecast results for each asset over 51 time periods (which means for example $51 \times 807 = 41,157$ forecasts just for each p in each model for the MSCI DM data set for example) we cannot report pseudo-R^2 values for every forecast individually. Instead we report for each data set the average pseudo-R^2 of all assets and all 51 time periods. The results presented in Table 2.3 show that the pseudo-R^2 values are rather low for all models, percentiles, and data sets. However, this is also the case for the models shown in Chevapatrakul (2013) and we are still able to gain several important insights from the pseudo-R^2 results. First, for all data sets we have the largest predictive power for Model 4 and the smallest predictive power for Model 1. For the MSCI DM and the MSCI USA data sets the predictive power for Model 3 is clearly better than for Model 2. For the MSCI EM, Model 2 performs slightly better than Model 3. Thereby we can note that the unconditional skewness and the turbulence contribution are useful explanatory variables for percentile forecasts. Second, for all models and all data sets the pseudo-R^2 value increases when the percentile p decreases. This is in line with the intuition that a higher predictive power can be achieved for percentiles that are further away from the mean. Third, for all models and percentiles p besides percentile $p = 0\%$ for Model 2 and Model 4, we have the highest predictive power for the MSCI USA data set, the second highest predictive power for the MSCI DM data set, and the worst predictive power for the MSCI EM data set.

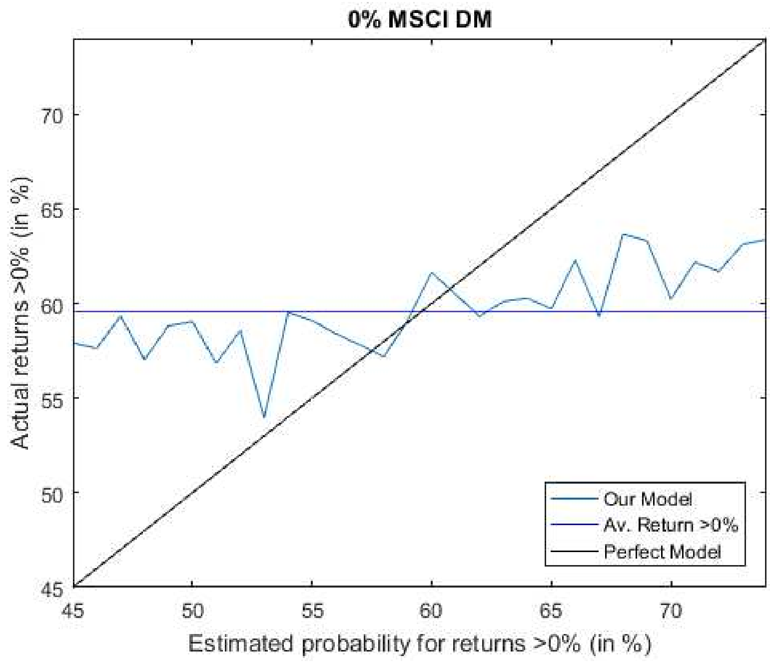
Table 2.4: Brier(Abs) Scores.

		0%	−2%	−4%	−6%	−8%
DM	Model 1	0.4797	0.4436	0.3966	0.3456	0.2956
	Model 2	0.4828	0.4492	0.4050	0.3558	0.3073
	Model 3	0.4755	0.4361	0.3864	0.3336	0.2830
	Model 4	0.4788	0.4415	0.3941	0.3427	0.2931
USA	Model 1	0.4714	0.4262	0.3705	0.3141	0.2587
	Model 2	0.4705	0.4269	0.3731	0.3185	0.2652
	Model 3	0.4636	0.4140	0.3557	0.2987	0.2434
	Model 4	0.4616	0.4137	0.3573	0.3022	0.2493
EM	Model 1	0.4905	0.4693	0.4391	0.4051	0.3675
	Model 2	0.4984	0.4848	0.4613	0.4319	0.3991
	Model 3	0.4810	0.4559	0.4215	0.3849	0.3453
	Model 4	0.4910	0.4736	0.4457	0.4131	0.3773

This table reports Brier(Abs) scores for five different values of p for the four models.

We explain in Section 2.4.1 that a few incorrect forecasts can have a large impact on the Brier(Sq) scores. Therefore we only report the scores of the Brier(Abs), which was developed by Christoffersen et al. (2007). Table 2.4 displays the Brier(Abs) values for each percentile p in each model for our three data sets. As for the pseudo-R^2 values, we report the average value of all assets and all 51 time periods. We can gain several important insights from the results displayed in Table 2.4. First, the Brier(Abs) scores are rather small for all data sets and percentiles, which indicates the good predictive power of our model. In particular it is noteworthy that all Brier(Abs) values are smaller than 0.5 and that rather high Brier(Abs) scores—such as for certain specifications of some of the data sets in Christoffersen et al. (2007)—do not occur. Second, the Brier(Abs) scores decrease (and therefore improve) when p decreases and we have excellent Brier(Abs) scores below 0.25 for some models and data sets for $p = -8\%$. Third, no model clearly dominates the others (or is dominated by all the other models), which was the case for the pseudo-R^2 measure. However, we have on average the smallest Brier(Abs) values for Model 3, and the largest Brier(Abs) values for Model 2. This implies that the turbulence contribution as explanatory variable has a positive effect on the quality of the forecast, while the effect of the unconditional skewness is negative. Fourth, we have the best Brier(Abs) scores for the MSCI USA data set and the worst Brier(Abs) scores for the MSCI EM data set. This indicates in combination with the pseudo-R^2 scores that forecasts for the MSCI EM data sets are particularly difficult. Intuitively this could be caused by the behavior of the emerging markets, which are more volatile in general than developed markets.

Figure 2.1: Model 1 Forecasting Performance Example.



This figure shows estimated probabilities for returns larger than 0% in comparison to average actual returns larger than 0% for Model 1 (MSCI DM).

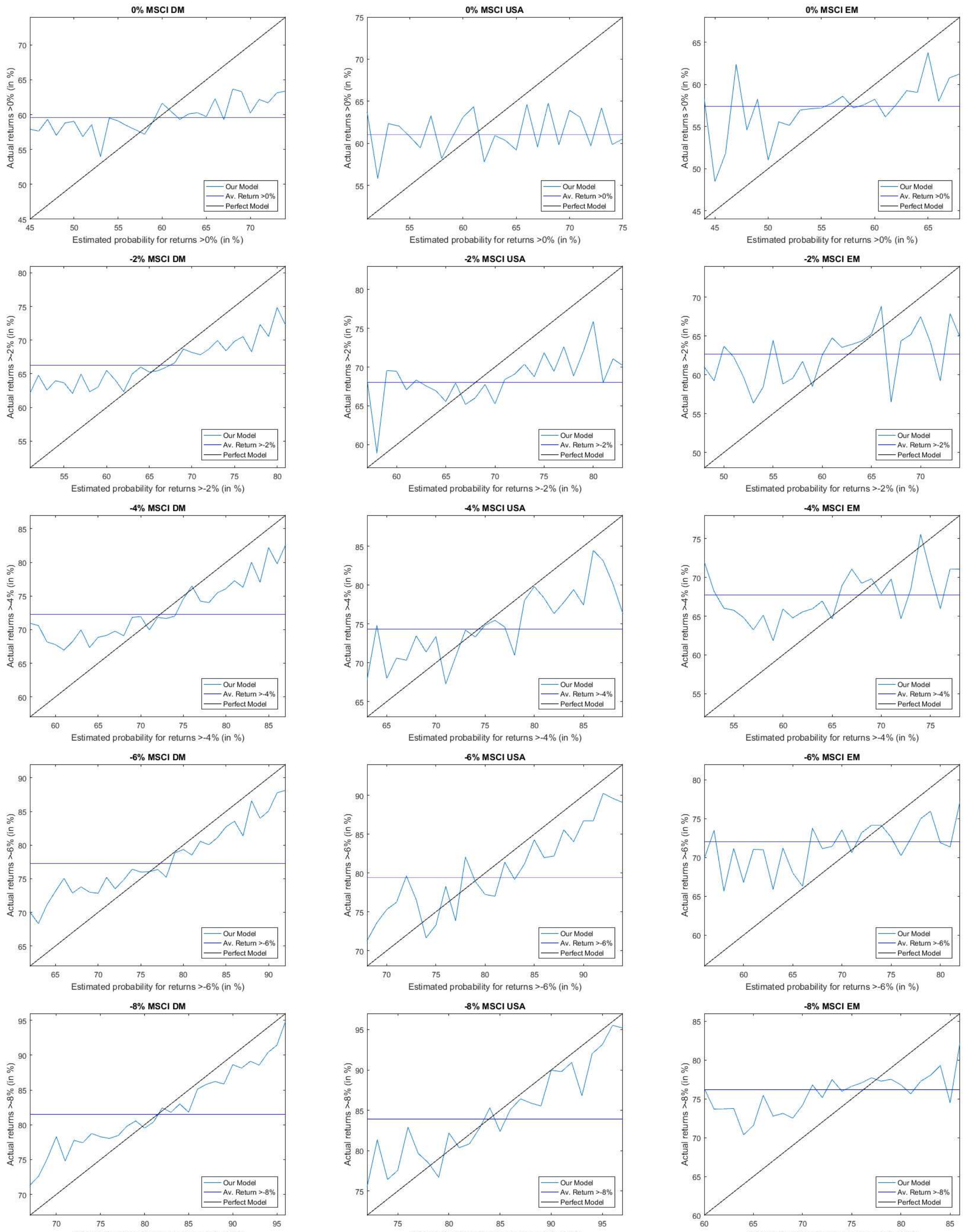
So far we only assessed the forecasting performance on an overall basis. However, we are also interested in the detailed forecasting accuracy for each of our estimated probability percentile buckets (which should not be confused with percentile p), which are described in Section 2.4.1, for example in how many percent of the cases that we predicted a return larger than p with a probability of 55 percent or 65 percent do we actually have a return larger than p. In a perfect model the estimated probability for a return larger than p should equal the average of the actual return larger than p, for example if the estimated probability for a return larger p is 55 percent or 65 percent the average actual return for estimated returns larger than p should be 55 percent or 65 percent, respectively. We analyze the performance of the estimated probability percentiles of our forecast with our own test, introduced in Section 2.4.1. This test is well suited for a visualization of the forecast results and we show the forecasting performance of Model 1 in Figure 2.2, the performance of Model 2 in Figure 2.3, the performance of Model 3 in Figure 2.4, and the performance of Model 4 in Figure 2.5. Since the plots in Figure 2.2–Figure 2.5 are rather small, we show the first plot in the first row of Figure 2.2—which represents Model 1 for $p = 0\%$ using the MSCI DM data set—separately in Figure 2.1 to explain the legends and the intuition behind our figures. On the x-axis in this and all other figures we only show the estimated probability percentile buckets that are represented in at least one percent of the total estimations. For example, for the MSCI DM data set we have $51 \times 807 = 41,157$ estimated probabilities in total, so we only show an estimated probability percentile if we have at least 412 estimations for this percentile. On the y-axis we show the average actual return for each of the estimated probability percentile buckets that are represented on the x-axis. The black $y = x$ line labelled “Perfect Model” represents a perfect model, in which the estimated probability larger than 0 percent exactly equals the average of the actual returns larger than 0 percent. Ideally the light blue line labelled “Our Model”—which represents the relationship between estimated probability percentile buckets and the average of their actual returns for our model—should resemble the black line of the ”Perfect Model”. As we can see in Figure 2.1 we are rather far away from this ideal. However, we can at least observe a non-monotonous increase of the average of actual returns larger than 0 percent when the estimated probability percentile bucket increases. This indicates that higher estimated probabilities for a return larger than 0 percent have a higher average of actual returns larger than 0 in general. The blue line labelled “Av. Return $> 0\%$” represents the average actual returns larger than p of all estimates. The graph of a poor model would only fluctuate around this line.

The plots in Figure 2.2–Figure 2.5 confirm some of the insights gained in the previous tests, but also provide several new findings. A comparison of the plots in our four figures supports the insight of the Brier(Abs) scores that the difference between the four models is rather small and that Model 3 has a slightly better forecasting performance than the other models. The three plots in the first row of Figure 2.2 (as do the first three plots of the three other figures) indicate that our model is far away from the perfect model for $p = 0\%$. However, a decrease of p leads to a better performance for all models, especially for the MSCI DM and the MSCI USA data sets. For example we can see from the second plot in the fifth row of Figure 2.4 that the forecast for Model 3 and $p = -8\%$ comes very close to the perfect model for the MSCI USA data set. The forecasting performance for the MSCI EM data set is much weaker than for the two other data sets. Besides the plot for Model 3 and $p = -8\%$ shown in the third column, fifth row of Figure 2.4 the plots for the MSCI EM data set are quite far away from the perfect model.

2.5 Trading Strategies Based on Percentile Forecasts

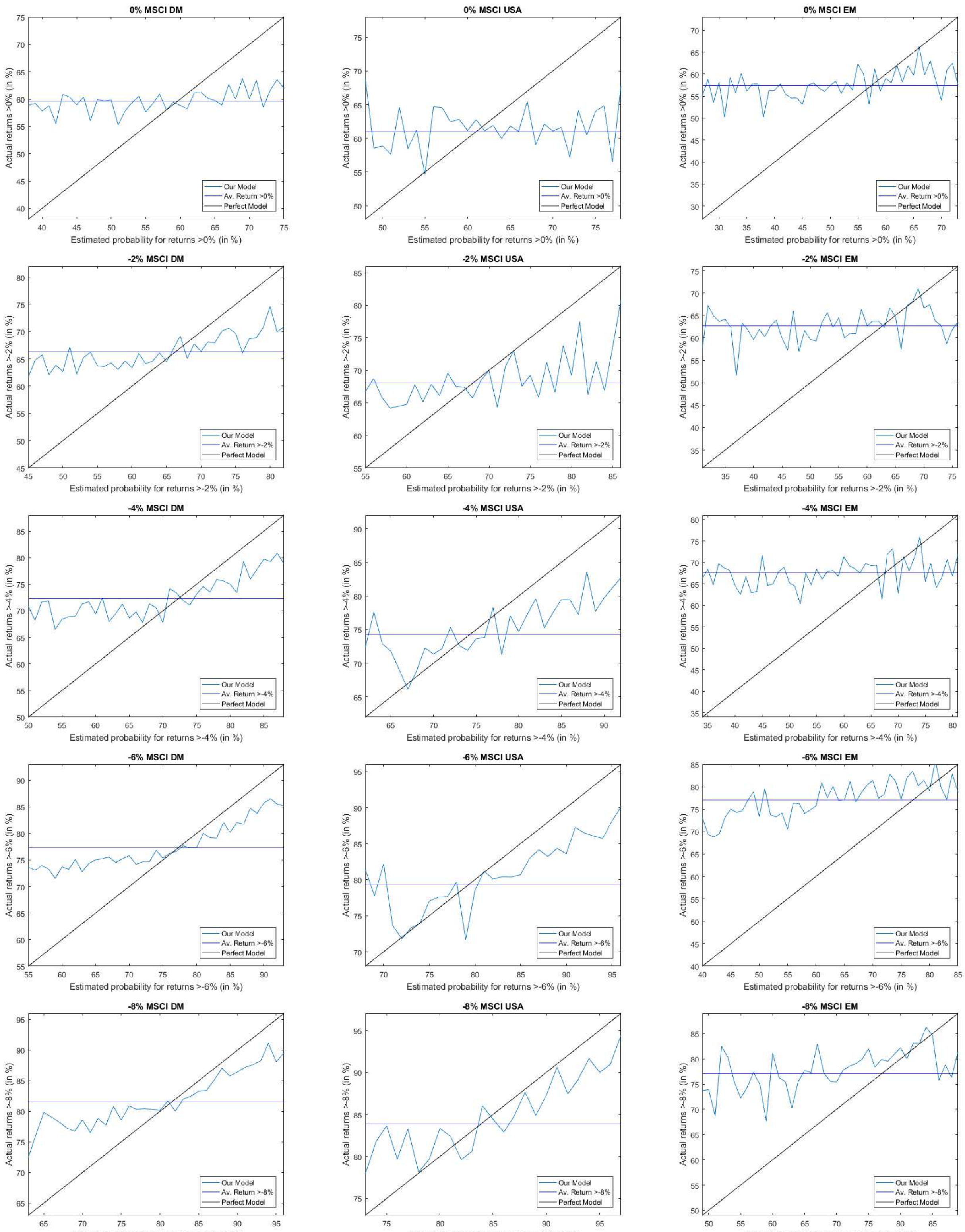
An essential question in the sign forecasting literature (see for example Christoffersen et al. (2007), Nyberg (2011), or Chevapatrakul (2013)) is how a high predictive forecasting power can be utilized in a profitable investment strategy. We are particularly interested in trading strategies with good performance and low risk; more precisely we are aiming for a strategy that provides high out-of-sample Sharpe ratios with low volatility. Furthermore we analyze if the higher predictive power that we found in Section 2.4 for smaller percentiles such as $p = -6\%$ or $p = -8\%$ compared to the traditional sign forecast predictions also leads to better performance in terms of Sharpe ratios and out-of-sample volatility. In Section 2.5.1 we explain our trading approach based on our percentile forecasts and introduce our performance measures briefly. In particular we develop a simple trading simulation in which the investor decides at the beginning of each quarter if he or she wants to invest in the stock market or in the risk-free asset based on the percentile forecasts. In Section 2.5.2 we present and analyze the performance of our trading strategy for our different models, data sets, and different values of p.

Figure 2.2: Model 1 Forecasting Performance.



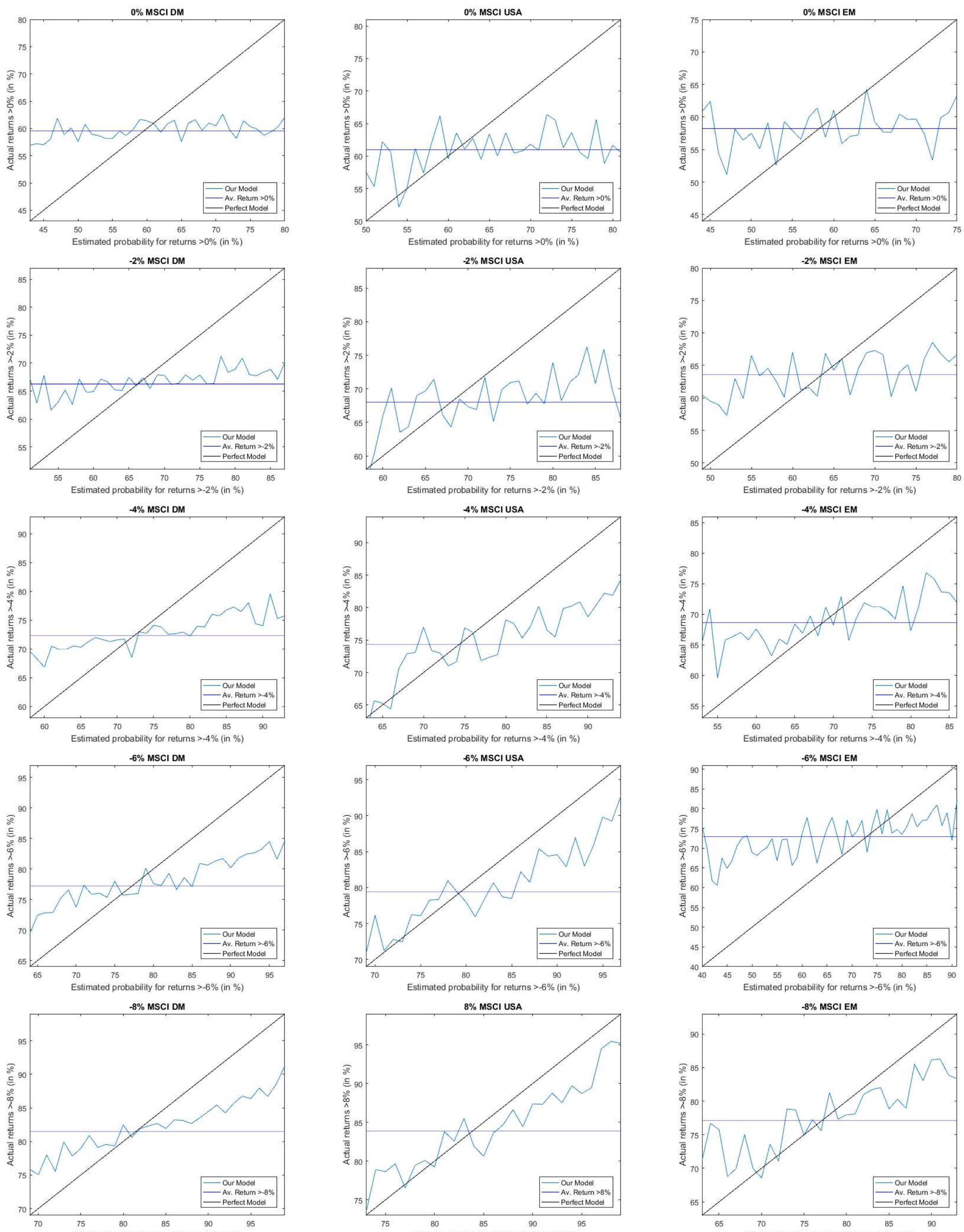
This figure shows estimated probabilities for returns larger than 0%, −2%, −4%, −6%, and −8% in comparison to average actual returns larger than 0%, −2%, −4%, −6%, and −8% for Model 1.

Figure 2.3: Model 2 Forecasting Performance.



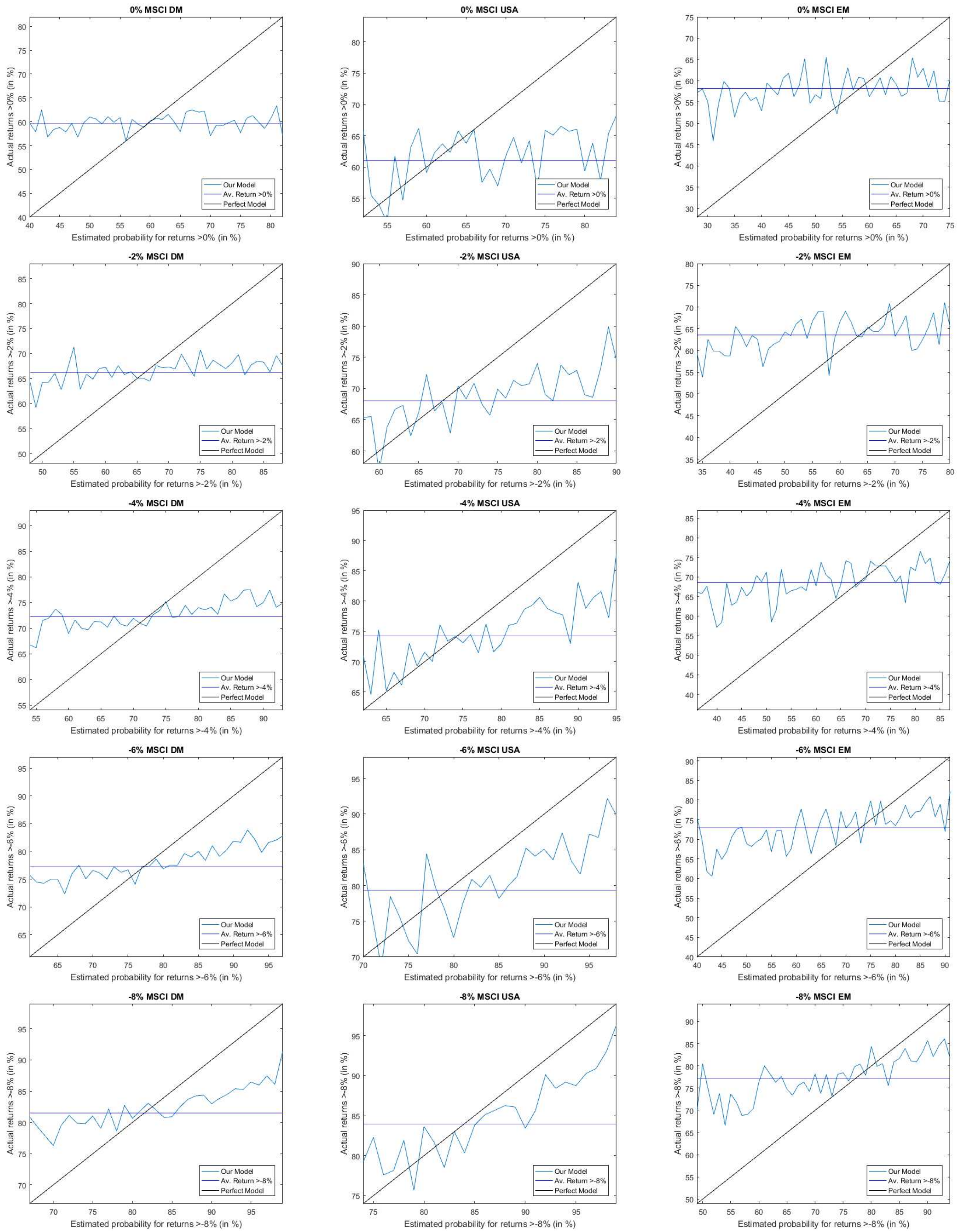
This figure shows estimated probabilities for returns larger than 0%, −2%, −4%, −6%, and −8% in comparison to average actual returns larger than 0%, −2%, −4%, −6%, and −8% for Model 2.

Figure 2.4: Model 3 Forecasting Performance.



This figure shows estimated probabilities for returns larger than 0%, −2%, −4%, −6%, and −8% in comparison to average actual returns larger than 0%, −2%, −4%, −6%, and −8% for Model 3.

Figure 2.5: Model 4 Forecasting Performance.



This figure shows estimated probabilities for returns larger than 0%, -2%, -4%, -6%, and -8% in comparison to average actual returns larger than 0%, -2%, -4%, -6%, and -8% for Model 4.

2.5.1 Trading Strategies

To assess the practical advantages of our percentile forecasts we develop a simple trading simulation, as do several other sign forecast studies, such as Pesaran and Timmermann (1995), Leung et al. (2000), Nyberg (2011), and Chevapatrakul (2013). At the beginning of every quarter an investor has two choices. He or she can invest in the stock market or in the risk-free rate.[8] The investor's investment decisions are guided by the percentile forecasts for each asset. We execute two calculations for the investor's decision. First, at each forecasting period k we calculate the percentage z of the realized returns of all assets n in all periods before the first forecasting period $(1, \ldots, B)$ and from the first forecasting period until k $(1, \ldots, k)$ that are larger than p:

$$z = \frac{1}{bn} \sum_{i=1}^{B} \sum_{j=1}^{n} I(R_{i,j} > p) + \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} I(R_{i,j} > p). \tag{2.23}$$

Second we calculate the mean m of the corresponding percentile probability forecasts for all assets n:

$$m = \frac{1}{n} \sum_{j=1}^{n} \widehat{Pr}_j(R_{k+1} > p|X_k). \tag{2.24}$$

If m is larger than a fraction γ of z we invest in the stock market, otherwise we invest in the risk-free asset. The choice of γ depends on the risk aversion of the investor. The more risk averse an investor is, the larger is the γ that investor would choose. A γ value larger than 1 would most likely result in investments in the risk-free rate more than 50 percent of the time. A γ value clearly smaller than 1 would most likely result almost exclusively in investments in the stock market. In Section 2.5.2 we analyze the performance of our trading strategies depending on γ and evaluate an appropriate choice for this parameter.

For the investment in the stock market we analyze the performance of two different strategies, which are simple, well-known, and have rather good out-of-sample performance: The $1/N$ portfolio and the global minimum variance (GMV) portfolio. The $1/N$ portfolio—which corresponds to the simple idea of investing an equal fraction of the total investment in each asset in the universe— is, according to De Miguel et al. (2009), not consistently outperformed out-of-sample by 14 popular portfolio optimization models. The GMV portfolio is a special case in the well-known portfolio optimization framework developed by Markowitz (1952). The asset holdings in this portfolio are solely based on variance-covariance estimations, which are more accurate than the mean estimations from a series of returns (see Merton (1980)). However, despite its rather simple input requirements the GMV portfolio often outperforms portfolios based on more advanced models out-of-sample (see Jorion (2000)).

We calculate the GMV portfolio by minimizing the variance of the portfolio according to Markowitz

[8] We also implemented several variations of two different, more advanced trading strategies. In the first of these more advanced strategies we excluded only assets with low probabilities from the asset universe, but invested in the stock market in each period. In the second strategy we allocated weights to our assets based on their probabilities. Neither of these strategies performed well out-of-sample.

(1952), using n potential assets

$$\min_{w} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_i \Sigma_{ij} w_j), \tag{2.25}$$

where the decision variables w_i are the weights of the holdings in each of the i^{th} potential assets and Σ is the variance-covariance matrix of the potential assets. We face the constraint that the weights w_j have to sum up to 1

$$\sum_{j=1}^{n} w_j = 1. \tag{2.26}$$

Furthermore we limit ourselves—as do many asset management companies in practice—to portfolios without short sales and therefore introduce the following non-negativity constraint:

$$w_j \geq 0 \quad \forall\ j. \tag{2.27}$$

For the investments in the risk-free rate we assume a return of 0 in every period.[9]

We are in particular interested in a low-risk investment strategy, without jeopardizing decent returns at the same time. Therefore we focus on two performance measures to assess the quality of our trading strategy: the out-of-sample standard deviation and the out-of-sample Sharpe ratio of the portfolio. A small standard deviation implies a small portfolio risk, while a reasonably high Sharpe ratio guarantees that an improvement of the standard deviation does not come with the price of significantly smaller returns. We also calculate statistical p-values to check the significance of the differences in out-of-sample Sharpe ratios between the portfolios based on the trading strategy and the portfolios that are based on the respective stock market only strategies. Furthermore we calculate the turnover of our investment strategies and of the respective stock market only strategies.

We annualize our first performance indicator—the standard deviation of our quarterly out-of-sample returns—using the standard method of the industry. We use the following formula to calculate σ_{an}:

$$\sigma_{an} = \sigma_q * \sqrt{4}. \tag{2.28}$$

Our next performance indicator is the out-of-sample Sharpe ratio. Following Sharpe (1994), we compare our trading portfolios based on the $1/N$ and GMV strategy, and the regular $1/N$ and GMV portfolios, to a benchmark. The risk-free rate—which we assume to be zero in all time periods—serves as our benchmark. At each adjustment point we subtract the risk-free rate from the out-of-sample return of our portfolios. This provides us the excess returns D of the portfolio return R_P over the risk-free rate benchmark R_F

$$D = R_{P,t} - R_{F,t} \quad \forall\ t. \tag{2.29}$$

Then we calculate the mean and the standard deviation of the excess returns D. Finally we divide the mean by the standard deviation for each of the portfolios to find the out-of-sample Sharpe ratio

[9] This is a small simplification. In most of the analyzed periods the risk-free rate was actually zero, while it was slightly larger than zero in some periods. Therefore exact values of the interest rates would slightly improve the performance of our models compared to the respective benchmarks.

SR

$$SR = \frac{\bar{D}}{\sigma_D}. \tag{2.30}$$

We also test whether the Sharpe ratios of our trading strategy portfolios based on the $1/N$ and the GMV strategy are statistically distinguishable from the Sharpe ratios of the $1/N$ and the GMV portfolio, respectively. As do De Miguel et al. (2009), we therefore compute the p-value of the difference, using the Jobson and Korkie (1981) approach and the correction of Memmel (2003).[10]

A potential weakness of our investment strategy is that a lot of trading between the stock market and the risk-free rate might be required, which causes transaction costs. To assess the trading amount of our investment approaches we compute the average portfolio turnover TO of all trading periods. We consider three different scenarios: investments in the stock market in consecutive periods, investments in the risk-free rate in consecutive periods, and a change of investments from the stock market to the risk-free rate and vice versa. If we invest in the stock market in consecutive periods we compute the turnover in stock market periods TO_s as per De Miguel et al. (2009):

$$TO_s = \sum_{j=1}^{n} \left(\left| w_{j,t+1} - w_{j,t^+} \right| \right), \tag{2.32}$$

where $w_{j,t+1}$ is the portfolio weight in asset j at time $t+1$ after rebalancing and w_{j,t^+} is the portfolio weight before rebalancing at $t+1$ (see De Miguel et al. (2009)). If we invest in the risk-free rate in consecutive periods we have a turnover of 0. In periods during which we change our investment from the stock market to the risk-free rate and vice versa we have a turnover of 2. For our portfolios we report the average turnover between period 2 and period T. Using the average portfolio turnover as a measure provides us with a meaningful comparison to the turnover of the benchmark strategies. And—in contrast to simpler measures, such as the percentage of portfolio changes—the average turnover considers both the negative effect of a change of investments (from the stock market to the risk-free rate and vice versa) and the positive effect of consecutive investments in the risk-free rate.

2.5.2 Trading Strategy Performance

In a first step of our analysis we evaluate the performance of our trading strategy for different values of γ. In Table 2.5 we report Sharpe ratios, p-values, annual standard deviation, and turnover for the MSCI DM data set, Model 1, and $p = 0\%$ for 28 different values of γ between 0.91 and 1.045. We gain several valuable insights from these results. First, for small γ-values (0.91 and 0.915) we invest almost exclusively in the stock market and our trading strategy performances converge to the respective benchmarks. For large γ values (1.045), on the other hand, we invest almost exclusively

[10] For our out-of-sample returns of the two portfolios i and n we obtain the test of the hypothesis $H_0 : \hat{\mu}_i/\hat{\sigma}_i - \hat{\mu}_n/\hat{\sigma}_n = 0$ using the test statistic $\hat{z}_{JK}$, which is asymptotically distributed as a standard normal

$$\hat{z}_{JK} = \frac{\hat{\sigma}_n \hat{\mu}_i - \hat{\sigma}_i \hat{\mu}_n}{\sqrt{\hat{\vartheta}}}, \; with \; \hat{\vartheta} = \frac{1}{A} \left(2\hat{\sigma}_i^2 \hat{\sigma}_n^2 - 2\hat{\sigma}_i \hat{\sigma}_n \hat{\sigma}_{i,n} + \frac{1}{2}\hat{\mu}_i^2 \hat{\sigma}_n^2 + \frac{1}{2}\hat{\mu}_n^2 \hat{\sigma}_i^2 - \frac{\hat{\mu}_i \hat{\mu}_n}{\hat{\sigma}_i \hat{\sigma}_n} \hat{\sigma}_{i,n}^2 \right), \tag{2.31}$$

where $\hat{\mu}_i$, $\hat{\mu}_n$, $\hat{\sigma}_i$, $\hat{\sigma}_n$, and $\hat{\sigma}_{i,n}$ are the estimated means, variances, and covariances and A is the number of adjustments (see De Miguel et al. (2009)).

Table 2.5: Investment Strategy Results for Different Values of γ.

	$1/N$				GMV			
γ	SR	p-value	σ_{an}	TO	SR	p-value	σ_{an}	TO
0.91	0.6847	0.44	0.1294	0.1572	1.2971	0.42	0.0783	0.2559
0.915	0.6847	0.44	0.1294	0.1572	1.2971	0.42	0.0783	0.2559
0.92	1.0024	0.10	0.1068	0.1544	1.6443	0.16	0.0674	0.2486
0.925	1.0024	0.10	0.1068	0.1544	1.6443	0.16	0.0674	0.2486
0.93	1.0024	0.10	0.1068	0.1544	1.6443	0.16	0.0674	0.2486
0.935	1.0024	0.10	0.1068	0.1544	1.6443	0.16	0.0674	0.2486
0.94	1.0024	0.10	0.1068	0.1544	1.6443	0.16	0.0674	0.2486
0.945	0.9633	0.13	0.1068	0.1524	1.5965	0.19	0.0678	0.2446
0.95	0.9633	0.13	0.1068	0.1524	1.5965	0.19	0.0678	0.2446
0.955	0.9633	0.13	0.1068	0.1524	1.5965	0.19	0.0678	0.2446
0.96	1.0164	0.12	0.1050	0.2279	1.6835	0.14	0.0659	0.3160
0.965	0.9805	0.14	0.1051	0.2254	1.6327	0.16	0.0664	0.3098
0.97	0.9418	0.17	0.1050	0.3021	1.5724	0.21	0.0665	0.3849
0.975	0.9354	0.17	0.0997	0.2989	1.4838	0.27	0.0672	0.3781
0.98	0.9354	0.17	0.0997	0.2989	1.4838	0.27	0.0672	0.3781
0.985	0.9423	0.16	0.0995	0.2974	1.4745	0.27	0.0673	0.3736
0.99	0.9360	0.17	0.0996	0.2950	1.5592	0.21	0.0654	0.3681
0.995	0.9360	0.17	0.0996	0.2950	1.5592	0.21	0.0654	0.3681
1	0.8811	0.21	0.0982	0.3720	1.4975	0.26	0.0640	0.4408
1.005	0.8811	0.21	0.0982	0.3720	1.4975	0.26	0.0640	0.4408
1.01	1.0168	0.13	0.0925	0.3702	1.5085	0.25	0.0638	0.4364
1.015	1.0985	0.09	0.0898	0.3685	1.5659	0.21	0.0627	0.4334
1.02	0.9785	0.15	0.0850	0.3654	1.4669	0.28	0.0629	0.4241
1.025	0.9920	0.20	0.0843	0.3611	1.4183	0.32	0.0633	0.4112
1.03	0.7767	0.37	0.0692	0.5481	1.1263	0.42	0.0565	0.5845
1.035	0.7390	0.40	0.0658	0.5001	1.0294	0.34	0.0536	0.5251
1.04	0.7311	0.41	0.0416	0.4872	0.7948	0.16	0.0433	0.4985
1.045	0.1783	0.12	0.0078	0.1612	0.4711	0.04	0.0162	0.1621
Benchmark	0.6407	-	0.1895	0.0905	1.2099	-	0.0878	0.2011

This table reports the out-of-sample Sharpe ratio (SR), annual standard deviation (σ_{an}), and the turnover (TO) for the investment strategy based on Model 1 and $p = 0\%$ for 28 different values of γ between 0.91 abd 1.045. The columns under the headings "p-value" show the statistical p-values of the difference between the Sharpe ratio of the investment strategy and the Sharpe ratios of the $1/N$ portfolio and the GMV portfolio, respectively. The last row reports the results for the $1/N$ portfolio and the GMV portfolio, which serve as benchmarks (MSCI DM).

in the risk-free rate and our trading strategy performs poorly. Second, for both the $1/N$ and the GMV approach the trading strategies outperform their respective benchmark for all γ values between 0.91 and 1.025. Third, the better performance, in terms of Sharpe ratios, of the trading strategy compared to the benchmark is slightly more significant for the $1/N$ than for the GMV strategy. Fourth, for the $1/N$ trading strategy the improvement in Sharpe ratios is highly based on a strong improvement of σ_{an}, which implies that the trading strategy is clearly less risky than the $1/N$ benchmark. For the GMV approach the improvement in σ_{an} is considerably smaller than for the $1/N$ approach. However, the σ_{an} of the GMV trading strategy is also much smaller than that of its benchmark (below 0.07 for every γ value larger than 0.915 compared to 0.0878 for the benchmark strategy). The difference between the two approaches is based on three periods with extreme return magnitudes (-18%, -24%, and 46%) that cause a high volatility of the $1/N$ benchmark approach. For most of the γ values our trading strategy leads to investments in the risk-free rate in these three periods. And a large share of the clearly smaller volatility of the trading strategy is based on the different investments in these three periods. The returns of the GMV benchmark approach are far less extreme than those of the $1/N$ benchmark and are in every period between -12 percent and 11 percent. Therefore there is less potential for our trading strategy to reduce σ_{an} compared to the GMV portfolio benchmark. Fifth, the trading strategies have a higher turnover than the benchmark strategy, and the improvements in Sharpe ratios of our trading strategies would partially disappear with the introduction of trading costs. However—especially for the smaller γ values—the differences in turnover between trading strategy and benchmark are still rather small. And the trading strategy would most likely outperform the benchmark in the presence of trading costs. Sixth, the trading strategy's performance does not increase or decrease monotonously with γ and several different values of γ are appropriate for an investor, as long as they are neither extremely small nor extremely large. A safe choice for γ should be considerably larger than 0.915 and considerably smaller than 1.03. For the remainder of this section we therefore calculate the 10 portfolios with γ values between 0.955 and 1 for our three data sets, four models, and five different values of p. To guarantee a clear presentation of our results we calculate each of the ten portfolios individually, but only report the average results of the 10 γ values for each p and each model in each data set.

In Table 2.6 we report the trading strategy performance results for the four models and the five values of p for the MSCI DM data set. These results support several insights gained from the analysis of the results in Table 2.5. First, the investments based on the trading strategy outperform the respective benchmarks for each model and each p. The difference in Sharpe ratios is more significant for the $1/N$ strategy than for the GMV strategy compared to their respective benchmarks. Second, σ_{an} is clearly smaller for the trading strategies than for the benchmark for each model and each p. Thereby, the difference in σ_{an} compared to the benchmark is much larger for the $1/N$ strategy than for the GMV strategy. Third, turnover increases for the trading strategies compared to their benchmarks for every p and every model. However, the increase is rather small. An exception is Model 2, which has a higher turnover for each p value than the three other models.

The results in Table 2.6 also provide several additional insights. First, the results are rather

Table 2.6: MSCI DM Investment Results.

		1/N				GMV			
Model	p	SR	p-value	σ_{an}	TO	SR	p-value	σ_{an}	TO
1	0%	0.9468	0.16	0.1018	0.2765	1.5543	0.22	0.0663	0.3562
	−2%	0.9562	0.16	0.1016	0.2608	1.5566	0.22	0.0663	0.3407
	−4%	0.9655	0.14	0.1018	0.2377	1.5702	0.21	0.0665	0.3182
	−6%	0.9682	0.14	0.1025	0.2234	1.5739	0.21	0.0667	0.3059
	−8%	0.9777	0.14	0.1031	0.2009	1.5740	0.21	0.0671	0.2845
2	0%	0.8886	0.26	0.0883	0.3808	1.4033	0.27	0.0616	0.4393
	−2%	0.9253	0.22	0.0930	0.3773	1.4629	0.28	0.0637	0.4420
	−4%	0.9466	0.18	0.0966	0.3410	1.5092	0.25	0.0647	0.4113
	−6%	0.9417	0.17	0.0979	0.3265	1.5132	0.25	0.0654	0.3989
	−8%	0.9476	0.17	0.0990	0.2963	1.5277	0.24	0.0656	0.3707
3	0%	1.0436	0.09	0.1040	0.1951	1.5982	0.19	0.0669	0.2811
	−2%	1.0505	0.08	0.1039	0.1792	1.6052	0.19	0.0671	0.2654
	−4%	1.0505	0.08	0.1039	0.1792	1.6052	0.19	0.0671	0.2654
	−6%	1.0449	0.09	0.1041	0.1872	1.5968	0.19	0.0673	0.2734
	−8%	1.0434	0.09	0.1043	0.1954	1.5932	0.19	0.0675	0.2818
4	0%	1.0388	0.10	0.0983	0.2807	1.4920	0.26	0.0635	0.3565
	−2%	1.0507	0.09	0.0991	0.2507	1.5167	0.25	0.0643	0.3284
	−4%	1.0536	0.09	0.0998	0.2282	1.5375	0.23	0.0646	0.3073
	−6%	1.0573	0.09	0.1005	0.2136	1.5575	0.22	0.0650	0.2938
	−8%	1.0612	0.09	0.1011	0.1909	1.5773	0.21	0.0654	0.2723
Benchmark		0.6407	-	0.1895	0.0905	1.2099	-	0.0878	0.2011

This table reports the out-of-sample Sharpe ratio (SR), annual standard deviation (σ_{an}), and the turnover (TO) for the investment strategy for the four models for five different values of p. The displayed results are the average for 10 different values of γ between 0.955 and 1. The columns under the headings "p-value" show the statistical p-values of the difference between the Sharpe ratio of the investment strategy and the Sharpe ratios of the $1/N$ portfolio and the GMV portfolio, respectively. The last row reports the results for the $1/N$ portfolio and the GMV portfolio, which serve as benchmarks (MSCI DM).

similar for all the models. However, the performance of Model 3 (for both strategies) and Model 4 (for the $1/N$ strategy) are slightly better than the performance of Model 1 and Model 2, which is in line with the results of the pseudo-R^2 values and Brier(Abs) scores shown in Section 2.4.2. Second, within each model the Sharpe ratios and σ_{an} are very similar for every value of p. However, on average over all four models the Sharpe ratios are larger for $p = -8\%$ than for $p = 0\%$ for both the $1/N$ strategy (1.0075 vs. 0.9795) and the GMV strategy (1.5681 vs. 1.5120). This is in line with the superior performance of smaller p-values in terms of pseudo-R^2 values and Brier(Abs) scores shown in Section 2.4.2. Additionally, the trading strategies based on $p = -8\%$ have for all models and both strategies (apart from Model 3 in the GMV strategy setup) a clearly smaller turnover than the strategies based on $p = 0\%$. Therefore this can be regarded as a small indicator that percentile forecasts are not only superior to simple sign forecasts in terms of forecast evaluation tests but also for trading strategies.

In Table 2.7 we report the trading strategy performance results for the MSCI USA data set. These results support several of our findings for the MSCI DM data set. The trading strategy outperforms the $1/N$ and the GMV benchmark, respectively, for all models and all values of p and the difference in Sharpe ratios is more significant for the $1/N$ strategy (than for the GMV strategy). Furthermore, Model 3 and Model 4 show a better performance than Model 1 and Model 2. This superior performance occurs for both the $1/N$ and the GMV strategy. However, there is also one major difference in the results between the two data sets. For the MSCI USA data set the Sharpe ratios for negative values of p are on average for all models and strategies smaller than for $p = 0\%$; that is to say, sign forecasting outperforms percentile forecasting in terms of Sharpe ratios. This is in contrast to the findings for the MSCI DM data set and also surprising in light of the superior test results for percentile forecasts in Section 2.4.2.

In Table 2.8 we report the trading strategy performance results for the MSCI EM data set. In contrast to the two other data sets, we introduced an upper bound of 0.2 for the maximum holding in each asset for the GMV strategy since an optimization with unconstrained holdings led to extreme positions in the portfolio despite a lower bound of zero and 231 potential assets in the universe (this is a common problem of mean-variance portfolios (see He and Litterman (1999))) and therefore also to extreme returns. Furthermore our trading rule led to investments solely in the risk-free rate for Model 2; therefore we do not report results for this model. In general the trading strategy does not work as well for the MSCI EM data set as it does for the two other data sets. However, this is line with the worse test results for the MSCI EM data set shown in Section 2.4.2. An intuitive explanation might be that percentile forecasts for emerging markets are less reliable than for developed markets. The trading strategy results still support several findings that we made for the two other data sets. First, Model 3 and Model 4 have a better performance than Model 1 in terms of Sharpe ratios. Second, with the exception of $p = 0\%$ for Model 4, the trading strategy outperforms the benchmark for all models and values of p for the $1/N$ approach. For the GMV approach the trading strategy also has better performance than the benchmark for most p values of Model 3 and Model 4. For all other models and p values—with the exception of $p = 0\%$ for Model 4—the Sharpe ratios are only

Table 2.7: MSCI USA Investment Results.

		1/N				GMV			
Model	p	SR	p-value	σ_{an}	TO	SR	p-value	σ_{an}	TO
1	0%	0.8892	0.18	0.1027	0.1931	1.4105	0.18	0.0696	0.2476
	-2%	0.8839	0.18	0.1033	0.1857	1.4168	0.18	0.0696	0.2406
	-4%	0.8916	0.17	0.1036	0.1706	1.4250	0.17	0.0695	0.2259
	-6%	0.8694	0.19	0.1063	0.1560	1.3956	0.20	0.0711	0.2123
	-8%	0.8799	0.18	0.1066	0.1484	1.4029	0.19	0.0711	0.2051
2	0%	0.8996	0.17	0.1030	0.1855	1.4252	0.17	0.0695	0.2403
	-2%	0.8931	0.17	0.1028	0.1924	1.4184	0.17	0.0695	0.2466
	-4%	0.9025	0.17	0.1025	0.1846	1.4102	0.18	0.0696	0.2389
	-6%	0.9009	0.17	0.1032	0.1701	1.4180	0.17	0.0696	0.2251
	-8%	0.9087	0.16	0.1029	0.1621	1.4147	0.18	0.0696	0.2171
3	0%	0.9907	0.11	0.1065	0.2888	1.5183	0.12	0.0706	0.3433
	-2%	0.9560	0.13	0.1064	0.2653	1.4934	0.13	0.0701	0.3212
	-4%	0.9347	0.13	0.1065	0.2501	1.4799	0.14	0.0700	0.3068
	-6%	0.9407	0.12	0.1071	0.2352	1.4751	0.14	0.0705	0.2926
	-8%	0.9407	0.12	0.1071	0.2352	1.4751	0.14	0.0705	0.2926
4	0%	1.0493	0.07	0.1082	0.3506	1.5288	0.12	0.0725	0.4027
	-2%	1.0223	0.09	0.1069	0.3421	1.5196	0.12	0.0713	0.3945
	-4%	0.9964	0.10	0.1063	0.3189	1.5067	0.13	0.0706	0.3723
	-6%	0.9695	0.12	0.1075	0.2809	1.4926	0.13	0.0710	0.3365
	-8%	0.9497	0.12	0.1075	0.2730	1.4834	0.14	0.0707	0.3293
Benchmark		0.6101	-	0.1840	0.0780	1.0125	-	0.1009	0.1448

This table reports the out-of-sample Sharpe ratio (SR), annual standard deviation (σ_{an}), and the turnover (TO) for the investment strategy for the four models for five different values of p. The displayed results are the average for 10 different values of γ between 0.955 and 1. The columns under the headings "p-value" show the statistical p-values of the difference between the Sharpe ratio of the investment strategy and the Sharpe ratios of the $1/N$ portfolio and the GMV portfolio, respectively. The last row reports the results for the $1/N$ portfolio and the GMV portfolio, which serve as benchmarks (MSCI USA).

Table 2.8: MSCI EM Investment Results.

		$1/N$				GMV			
Model	p	SR	p-value	σ_{an}	TO	SR	p-value	σ_{an}	TO
1	0%	0.7862	0.32	0.1845	0.2009	0.9254	0.42	0.1355	0.2446
	−2%	0.8064	0.34	0.1826	0.2005	0.9384	0.42	0.1348	0.2437
	−4%	0.8064	0.34	0.1826	0.2005	0.9384	0.42	0.1348	0.2437
	−6%	0.8125	0.34	0.1831	0.2008	0.9444	0.42	0.1352	0.2441
	−8%	0.7867	0.32	0.1854	0.1937	0.9274	0.41	0.1361	0.2379
3	0%	0.8846	0.31	0.2024	0.2031	1.0020	0.46	0.1393	0.2472
	−2%	0.8977	0.29	0.1878	0.2018	0.9995	0.45	0.1361	0.2450
	−4%	0.8902	0.30	0.1815	0.2015	0.9910	0.45	0.1348	0.2445
	−6%	0.8661	0.33	0.1836	0.1945	0.9791	0.45	0.1356	0.2384
	−8%	0.8579	0.34	0.1840	0.1799	0.9838	0.45	0.1355	0.2247
4	0%	0.6629	0.31	0.1009	0.5301	0.8081	0.33	0.0801	0.5406
	−2%	0.8282	0.32	0.1155	0.5322	0.9414	0.30	0.0912	0.5469
	−4%	0.9443	0.30	0.1365	0.5298	1.0236	0.24	0.1045	0.5503
	−6%	0.9958	0.23	0.1490	0.4736	1.0295	0.22	0.1145	0.4992
	−8%	1.0004	0.23	0.1568	0.4212	1.0288	0.22	0.1205	0.4503
Benchmark		0.7419	-	0.2587	0.1128	0.9875	-	0.1511	0.1663

This table reports the out-of-sample Sharpe ratio (SR), annual standard deviation (σ_{an}), and the turnover (TO) for the investment strategy for the four models for five different values of p. The displayed results are the average for 10 different values of γ between 0.955 and 1. The columns under the headings "p-value" show the statistical p-values of the difference between the Sharpe ratio of the investment strategy and the Sharpe ratios of the $1/N$ portfolio and the GMV portfolio, respectively. The last row reports the results for the $1/N$ portfolio and the GMV portfolio, which serve as benchmarks. The chosen γ values lead to investments solely in the risk-free rate for Model 2 (MSCI EM).

slightly smaller than those of the benchmarks. Third, the results for the MSCI EM data set support the insight that no percentile forecast is consistently superior to the others. For Model 3, we have the best performance for $p = 0\%$ and the worst for $p = -8\%$, while this is reversed for Model 4. However, it is noteworthy that we have by far the worst performance for $p = 0\%$ in Model 4.

Combining the results and interpretation of the three data sets we can record the following main insights. First, our trading strategy based on percentile forecasting leads in general to better performance in terms of out-of-sample Sharpe ratio and out-of-sample standard deviation compared to the $1/N$ and the GMV portfolio, respectively. This better performance is more significant in the comparison to the $1/N$ portfolio. Second, the trading strategies have a larger turnover than the $1/N$ and the GMV portfolios, respectively. This would reduce the superior performance slightly in the presence of transactions costs. Third, in line with the test results presented in Section 2.4.2 the investments based on Model 3 or Model 4 have in general higher Sharpe ratios than those based on Model 1 or Model 2. This shows that the turbulence contribution is well suited as an explanatory variable for percentile forecasts. Fourth, despite the clearly superior test performance presented in Section 2.4.2, strategies based on forecasts for smaller percentiles such as $p = -8\%$ or $p = -6\%$ do not consistently outperform the trading strategies based on sign forecasts. Fifth, our trading strategies based on percentile forecasts work very well for developed markets. However, for emerging markets too, investments based on our trading strategies do not lead to significantly worse Sharpe ratios than the respective benchmark strategies.

2.6 Conclusion

In this paper we have developed a binary logit approach to predicting the signs of individual asset returns. In a next step we extended this approach to a model that also forecasts the probability of returns larger than several percentiles. In particular we predicted the probabilities for returns larger than $p = 0\%$ (which equals the sign forecast), $p = -2\%$, $p = -4\%$, $p = -6\%$, and $p = -8\%$.

Our binary logit approach includes four different models that differ with regard to their explanatory variables. The first model uses only conditional volatility as an explanatory variable, the second one adds unconditional skewness, the third includes—as a novelty in the sign forecasting literature—in addition to the conditional volatility the turbulence contribution based on the Mahalanobis distance, which is a measure of financial turbulence (see Kritzman and Yuanzhen (2010)), and the fourth uses the conditional volatility, the unconditional skewness, and the turbulence contribution.

We introduced several tests for an ex post evaluation of our forecasts and implemented the models using three real-life data sets. The results of our tests demonstrated that prediction accuracy increases for the smaller percentiles. Furthermore, the tests showed that Model 3 and Model 4—which both use the turbulence contribution as an explanatory variable—provide a higher predictive power than Model 1 and Model 2.

We also developed a trading strategy based on our percentile forecasts. This strategy leads in general to better performance in terms of Sharpe ratios than the respective benchmark strategies with solely stock market investments. In line with superior forecast evaluation test results, the trading strategies based on Model 3 or Model 4 also lead to larger Sharpe ratios than do those based on Model 1 or Model 2. However, the superior test results for smaller percentiles ($p = -2\%$, $p = -4\%$, $p = -6\%$, and $p = -8\%$) compared to the sign forecast ($p = 0\%$) could not be transformed into consistently higher Sharpe ratios in our trading strategy.

2.A The Normal Inverse Gaussian Distribution

The NIG distribution is a subclass of the family of generalized hyperbolic (GH) distributions $GH(\mu, \delta, \alpha, \beta, \lambda)$ with fixed value $\lambda = -1/2$. The remaining parameters μ, δ, α, and β define location, scale, skew, and shape respectively.

The density of a NIG-distributed random variable X has the following form (see for example Barndorff-Nielsen and Blaesild (1981) and Eberlein and Prause (2002)):

$$f_{NIG}(x; \mu, \delta, \alpha, \beta) = \frac{\alpha\delta}{\pi} \frac{K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\sqrt{\delta^2 + (x-\mu)^2}} e^{\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)\right)}, \tag{2.33}$$

where K_1 is the modified Bessel function of the third kind

$$K_1(y) = \frac{1}{2}\int_0^\infty \exp\{-\frac{y}{2}(y + y^{-1})\}dy. \tag{2.34}$$

The conditions for a viable density are $\mu \in R$, $\delta > 0$, $\alpha > 0$, and $|\beta|/\alpha < 1$.

The NIG distribution has the desirable property of being invariant to transformations in location and scale; that is to say, if we standardize a NIG-distributed random variable X by its mean μ and variance σ, the resulting standardized variable $Z = (X - \mu))/\sigma$ will also be NIG distributed with unchanged skew and shape parameters. Employing a scale and location invariant parametrization $f_{NIG}(\zeta, \rho)$ it can be shown that[11]

$$f_{NIG}(x; \mu, \sigma, \rho, \zeta) = \frac{1}{\sigma} f_{NIG}(z; 0, 1, \rho, \zeta). \tag{2.35}$$

The new skew and shape parameters

$$\begin{aligned} \zeta &= \delta\sqrt{\alpha^2 - \beta^2} \\ \rho &= \frac{\beta}{\alpha} \end{aligned} \tag{2.36}$$

jointly determine skewness and kurtosis.

The steps from the (ρ, ζ) representation to the $(\mu, \delta, \alpha, \beta)$ parametrization, while at the same time standardizing for zero mean and unit variance, are explained in detail in the appendix of Ghalanos et al. (2015)).

The mean, variance, skewness, and kurtosis of a random variable $X \sim NIG(\mu, \delta, \alpha, \beta)$ are

$$\begin{aligned} E(X) &= \mu + \delta\frac{\beta}{\gamma} \\ V(X) &= \delta\frac{\alpha^2}{\gamma^3} \\ S(X) &= 3\frac{\beta}{\alpha\sqrt{\delta\gamma}} \\ K(X) &= 3 + 3\left(1 + 4\left(\frac{\beta}{\alpha}\right)^2\right)\frac{1}{\delta\gamma} \end{aligned}$$

[11] See Ghalanos et al. (2015) for a proof.

where $\gamma = \sqrt{\alpha^2 - \beta^2}$.

2.B Asymmetric GARCH

The GJR GARCH model of Glosten et al. (1993) differentiates between positive and negative shocks on the conditional variance. The asymmetry is introduced with the aid of the indicator function I, taking values of 1 for $\epsilon_t \leq 0$ and 0 otherwise:

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \left(\alpha_j \epsilon_{t-j}^2 + \gamma_j I_{t-j} \epsilon_{t-j}^2\right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{2.37}$$

The new parameter γ_j is usually referred to as the leverage term.

Essay 3

Replicating Portfolios in the Life Insurance Industry

A Replicating Portfolio Optimization Model[1]

Maximilian Adelmann
Dept. of Business Administration
University of Zurich
Moussonstrasse 15
8044 Zurich, Switzerland
maximilian.adelmann@business.uzh.ch

June 2015

Abstract

Replicating portfolios have recently emerged as an important tool in the life insurance industry, used for the valuation of companies' liabilities. This paper presents a replicating portfolio (RP) model for approximating life insurance liabilities as closely as possible. We minimize the L_1 error between the discounted life insurance liability cash flows and the discounted RP cash flows over a multi-period time horizon for a broad range of different future economic scenarios. We apply two different linear reformulations of the L_1 problem to solve large-scale RP optimization problems and also present several out-of-sample tests for assessing the quality of RPs. A numerical application of our RP model to empirical data sets demonstrates that the model delivers RPs that match the liabilities rather closely. We complete the paper with a comparison of running times for the two linear formulations and for different LP algorithms. The numerical analysis demonstrates that our model delivers RPs with excellent practical properties in a reasonable amount of time.

Keywords: Liability cash flows; linear programming; out-of-sample tests; replicating portfolios.

Note: A later version of this paper has been submitted to Operations Research under the title "A Large-Scale Optimization Model for Replicating Portfolios in the Life Insurance Industry" .

[1]I am heavily indebted to my supervisors János Mayer and Karl Schmedders for their support and guidance on this project. Also, I thank Regine Scheder and Lucio Fernandez Arjona for helpful discussions on the subject and feedback on earlier implementations of the model. Finally, I am grateful to seminar audiences at the University of Zurich and EURO2015 in Glasgow for helpful comments.

3.1 Introduction

Replicating portfolios have recently emerged as an important tool in the life insurance industry, used for the valuation of companies' liabilities. This paper presents a replicating portfolio (RP) model for approximating life insurance liabilities as closely as possible. We minimize the L_1 error between the discounted life insurance liability cash flows and the discounted RP cash flows over a multi-period time horizon for a broad range of different future economic scenarios. We apply two different linear reformulations of the L_1 problem to solve large-scale RP optimization problems and also present several out-of-sample tests for assessing the quality of RPs. A numerical application of our RP model to empirical data sets demonstrates that the model delivers RPs that match the liabilities rather closely. We complete the paper with a comparison of running times for the two linear formulations and for different LP algorithms. The numerical analysis demonstrates that our model delivers RPs with excellent practical properties in a reasonable amount of time.

The Solvency II Directive is a new regulatory framework in the European Union; it aims to harmonize and modernize European insurance supervision. The European Parliament approved the Solvency II framework in April 2009. It came into effect on January 1, 2016. Similar to the objectives of the Basel II framework for the banking and finance industry, Solvency II sets out to establish new capital requirements, valuation techniques, and governance and reporting standards for the insurance industry. For example, the new capital requirements should ensure that an insurance company has a sufficient amount of capital to avoid bankruptcy within the forthcoming year with a confidence level of (at least) 99.5 percent (see, for example, Devineau and Chauvigny (2011)). The directive's regulatory requirements include certain quantitative requirements, such as the Solvency Capital Requirement (SCR). Solvency II allows insurance companies to build their own internal models to determine their SCR or to use a standard formula provided by the regulators. The computation of metrics such as SCR requires insurance companies to calculate the "fair" (or "market-consistent") values of their insurance liabilities (see Schrager (2008)). When these liabilities are long-term and dependent on market conditions—as are, for example, the values of life insurance policies—an accurate valuation of such liabilities quickly becomes very complex. Several different techniques for the (approximate) evaluation of life-insurance liabilities have become increasingly popular in recent years, among them approaches based on replicating portfolios.[2]

In the insurance context, a replicating portfolio is a pool of a finite number of selected financial instruments designed to (approximately) reproduce the cash flows or present values of liabilities across a large number of economic scenarios. The practical benefits of RPs include their versatility. Insurance companies can use them for asset liability management (ALM) and performance manage-

[2]Due to the complexity of life insurance liabilities, evaluation approaches based on Monte Carlo simulations appear to be an attractive tool. However, both Schrager (2008) and Natolski and Werner (2015) argue that such brute-force approaches require too many calculations for the results to be sufficiently accurate. Devineau and Chauvigny (2011) claim that an accelerated method, the nested simulation approach, is best suited to the Solvency II context but also concede that this approach requires many calculations. With the current state of computing power, simulation-based approaches are essentially useless in computational practice due to excessive computation times.

ment, for risk management, for capital and value calculations, and for management information; see Boekel et al. (2009). While the idea of portfolio replication has a long history in derivative pricing (see Fouque et al. (2000)) and was already fundamental to the Black–Scholes model (see Black and Scholes (1973)), its application to insurance problems appears to be fairly recent. To the best of our knowledge, Pelsser (2003) is the first published paper to suggest static replicating portfolios as a viable alternative for insurance companies to use in valuing their life insurance products.

For an insurance company to reap the practical benefits of an RP, two conditions must be satisfied. First, the RP must replicate the liability cash flows as closely as possible for a large number of scenarios. This condition requires a metric for the closeness of cash flows for such replication. Second, it must be possible to compute an RP fairly quickly. This condition requires the setting up of a tractable optimization problem that determines the best possible RP with respect to the chosen metric.

An RP optimization process for liability cash flows consists of three main steps. In the first step, the data sets are generated. Firstly, scenario sets for different economic future events are determined. Secondly, scenarios for the liabilities of the insurance company and for a set of candidate assets for the RP are simulated for each of these scenario sets. These simulations result in a vector of liability cash flows and vectors of asset cash flows for each asset. In the second step, first a metric for the distance between the vector of liability cash flows and the cash flows from a portfolio of candidate assets is chosen. Then the solution of an optimization problem minimizing the distance metric determines the asset weights in the RP. Finally, in the third step, the quality of the RP is examined in the light of several tests.

In this paper we focus on the second and third steps of the RP optimization process. We take a set of candidate assets and cash flow vectors for liabilities and assets as given. We then discuss the formulation of an optimization problem for determining an RP. In particular, we discuss in detail useful constraints that help us to find RPs with favorable properties. During the discussion of the third step, we propose several out-of-sample tests for the evaluation of RPs.

The simplest approach to the RP optimization—see Daul and Gutiérrez Vidal (2009)—would be to minimize the sum of squared errors, the L_2 norm, without any constraints because then the optimal portfolio weights are given by the standard least-squares analytical solution. However, it is well known that for the detection of outliers (see, for example, Bloomfield and Steiger (1983) and Bektas and Sisman (2010)) the L_1 norm is superior to the L_2 norm. Moreover, Rice and White (1964) show that for estimation the L_1 norm is the best choice if the error distribution has long tails. And so, since the input data sets generated by insurance companies likely include extreme cash flows, the L_1 norm appears to be the most appropriate choice for the objective function. In addition, since L_1 minimization problems allow for linear reformulations, the inclusion of linear constraints represents no challenge.

We employ two different linear reformulations of the L_1 minimization problem. The first reformulation—see Konno and Yamazaki (1991)—has been previously suggested for RPs, (see Chen and Skoglund (2012)). The second reformulation—see Feinstein and Thapa (1993)—has not yet appeared in the RP literature. Both linear reformulations enable us to solve the RP optimization problem for liability cash flows over many years. Furthermore, the linear formulations allow for a computationally cheap introduction of additional constraints. We employ two different sets of constraints. The first set is supposed to ensure a relatively small RP that includes only a small fraction of the candidate assets. The advantage of a small RP is that it is quicker to price and easier to interpret in relation to the liability cash flows compared to a big RP with many different asset positions (Burmeister and Mausser (2009)). The second set of constraints is supposed to ensure small in-sample scenario set movements. These small scenario set movements ensure that the RP will likely perform in the same way as the mapped liabilities in a large range of different future scenarios.

After the complete formulation of the linear RP optimization problems, we also present several tests for assessing the quality of the solutions to the RP optimization problems. The most important of these tests focus on out-of-sample scenario set movements and the market value fit between the liability cash flows and those of the RP. These tests, which are performed with out-of-sample data, validate that the RP is most likely able to behave in the same way as the liabilities for different economic scenarios. In particular, these tests are meant to show that solving the RP problem does not lead to overfitting for a specific in-sample data set. Another common test for assessing the RP's quality—see, for example Daul and Gutiérrez Vidal (2009) and Burmeister and Mausser (2009)—is the coefficient of determination, R^2, between RP and liability cash flows.

After the theoretical formulation of our RP optimization model and the subsequent tests, we numerically solve a large-scale RP optimization problem and test this solution's results. We use three real-life data sets, which we received from Zurich Insurance Group Ltd.[3] The first serves as in-sample input for our optimization model. It includes 13 scenario sets with 100 scenarios each and 919 candidate assets, and spans a period of 40 years. The two additional data sets serve as out-of-sample data for testing and assessing the quality of our optimization results. The first of these test data sets has exactly the same size as the optimization data set and is used to test the out-of-sample scenario set movements. The second test data set is used to test the market value fit of the RP. This large amount of available out-of-sample data for testing purposes is a great strength of our analysis. We solve the linear RP problems for various parameter values in the constraints and conduct the tests for their solutions. Doing so enables us to present a thorough sensitivity analysis of the optimal RPs and to examine their strengths and weaknesses. We solve the RP problems using the solver CPLEX® in GAMS®.[4] We report detailed running times for three different solvers in CPLEX for the two linear RP formulations.

[3]The Zurich Insurance Group Ltd. is a global insurance company headquartered in Zurich, Switzerland. It employs approximately 55,000 people worldwide and operates in more than 170 countries. It is a Swiss blue chip with a market value of around USD 35 billion; see `www.zurich.com/en/about-us/facts-and-figures` (last accessed October 26, 2015) for further information. In the remainder of this paper we refer to the company as "Zurich".

[4]CPLEX and GAMS are a registered trademarks. In the remainder of this paper we suppress all trademark signs. All our calculations in GAMS reported in this paper were performed with GAMS 24.2.2. For more information about GAMS, see Brooke et al. (2014).

The Literature

Replicating portfolios have a long history in derivative pricing. Since this literature is of little if any relevance to the applications of RPs in the insurance industry, we do not review this vast literature and just refer to Fouque et al. (2000) who provides an introduction to RPs in the context of the Black-Scholes model for derivative pricing. In this context RPs are used for dynamic hedging strategies. For the remainder of this literature review we focus on publications presenting and analyzing the concepts of static portfolio replication and RPs in the insurance industry.

Schrager (2008) describes RPs and emphasizes that they can be used as an important tool for risk-based solvency calculations. Koursaris (2011) provides an introduction to RPs. Boekel et al. (2009) demonstrate the use of RPs as a pool of assets to reproduce the cash flows of a pool of liabilities across a large number of stochastic scenarios. They point out such portfolios' great benefits for the recalculation of the effects of financial market developments and as a sophisticated tool for risk aggregation. Daul and Gutiérrez Vidal (2009) present various techniques for constructing an RP based on the L_2 norm for the cash flow matching. They analyze different approaches and study practical aspects using case studies indicating the robustness of the replication technique and presenting an effective way of evaluating the quality of the replication. Seemann (2009) explains the theory of RPs and the complete RP process at great length and provides a detailed numerical example. For the RP calibration process he suggests the L_2 norm. He points out the importance of the available set of candidate assets and presents an outlook for the use of RPs in the life insurance industry. Devineau and Chauvigny (2011) present common applications of the RP technique. They urge the practitioner to consider several factors for an optimization process, including the makeup of the portfolio, the parameters of its assets, the selection of the optimization program, and the type of variable that has to be replicated. They provide different calibration methods, which are all based on the L_2 norm. Oechslin et al. (2007) propose a terminal-value matching approach for RPs as a superior alternative to cash-flow matching models. Natolski and Werner (2014) compare different quadratic measures for RPs and introduce a new discounted terminal-value matching approach. They do not allow for constraints in the optimization problem whereas, we have constraints. Burmeister and Mausser (2009) stress the advantages of an RP with a small number of assets. They establish trading constraints in their optimization model to select the most relevant instruments to improve the performance of the RP. Burmeister et al. (2010) show that trading constraints are an effective way of regularizing optimization problems so that they produce small RPs. Based on several experiments they prove that trading costs based on simple statistics provide good performance with minimal computational effort. Burmeister and Mausser (2009) and Burmeister et al. (2010) both present the L_2 and the L_1 methods for the optimization process. However they do not mention the two linear formulations of the L_1 norm that we apply in this paper. Chen and Skoglund (2012) describe an efficient linear programming approach to cash flow replication. First they briefly show how cash flow mismatch can be used as an objection function, and they also mention the first of the two linear reformulations of the L_1 norm in a brief footnote. For the remainder of their paper they focus on cash flow mismatches as constraints. Their analysis is purely theoretical and they do not provide any case studies that use real-life data.

The remainder of this paper is organized as follows. We introduce the RP optimization models in Section 3.2. Section 3.3 presents our tests for assessing the quality of our optimization results. In Section 3.4 we describe the data sets we use and show the specific optimization problem solved in this paper. Section 3.5 compares the results of different RPs, provides more detailed results for three selected RPs, and reports running times. Section 3.6 concludes.

3.2 The RP Model

In this section we develop two linear RP optimization models.

3.2.1 The Basic RP Problem

The principal objective of a replicating portfolio is to reproduce given liability cash flows with a pool of asset cash flows for a broad range of economic scenarios. In mathematical terms, we basically want to solve the following system of linear equations:

$$\begin{pmatrix} ACF_{1,1} & ACF_{1,2} & \cdots & ACF_{1,A} \\ ACF_{2,1} & ACF_{2,2} & \cdots & ACF_{1,1} \\ \vdots & \vdots & \ddots & \vdots \\ ACF_{C,1} & \vdots & \cdots & ACF_{C,A} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_A \end{pmatrix} = \begin{pmatrix} LCF_1 \\ LCF_2 \\ \vdots \\ LCF_C \end{pmatrix} \tag{3.1}$$

where $ACF_{c,a}$ is the asset cash flow of asset a for scenario c and LCF_c is the liability cash flow for scenario c. The cash flows $ACF_{c,a}$ and LCF_c are the data for the replication problem. The unknown solution is the vector of portfolio weights x_a for the assets a.

For typical applications in the life insurance industry, the number of scenarios vastly exceeds the number of candidate assets. For example, our data set—provided by Zurich—includes 919 assets and a total of $52,000$ scenarios. Therefore, generically the problem (3.1) does not have a solution. So, instead of solving the system of linear equations (3.1) we try to match the RP cash flows on the left-hand side and the liability cash flows on the right-hand side as closely as possible. The simplest approach to determining the portfolio weights—see Daul and Gutiérrez Vidal (2009)—would be to minimize the sum of squared errors, the L_2 norm, between the RP cash flows and the liability cash flows. In this approach the optimal portfolio is given by the standard least-squares analytical solution. However, it is well known that for the detection of outliers (see, for example, Bloomfield and Steiger (1983) and Bektas and Sisman (2010)) the L_1 norm is superior to the L_2 norm. Moreover, Rice and White (1964) explain that "for the important problem of smoothing and estimation in the presence of wild points, the L_1 norm appears to be markedly superior among the L_p norms." Since insurance companies may be particularly interested in the performance of their insurance products under extreme economic conditions, the data sets may likely include "wild points". We therefore choose the L_1 norm as the objective function of our RP model. In addition, since L_1 minimization problems allow for linear reformulations, the inclusion of linear constraints in our model represents no computational challenge.

3.2.2 Objective

Before we define the objective function, we first refine the notation for our model. We try to replicate the discounted cash flows of life insurance liabilities using a set of different asset classes, $a = 1, \ldots, A$. Each of these asset classes contains a set of I_a, $a = 1, \ldots, A$, candidate assets. The discounted cash flows occur at discrete times $t = 1, \ldots, T$, where $t = 1$ is the first year and T is the final year of the planning horizon. In the model we face different sets of scenarios, $k = 1, \ldots, K$. The first set of scenarios, $k = 1$, is the so-called base scenario set, which is comprised of different scenarios in line with current market prices. The remaining scenario sets, $k = 2, \ldots, K$, contain scenarios that reflect changes in critical market factors. Every scenario set includes a set of simulated scenarios $c = 1, \ldots, C$. Table 3.1 summarizes the indices for our model.

Table 3.1: List of Indices.

Notation	Name
$k = 1, \ldots, K$	Scenario sets
$c = 1, \ldots, C$	Scenarios
$t = 1, \ldots, T$	Time points
$a = 1, \ldots, A$	Asset classes
$i = 1, \ldots, I_a$	Candidate assets

This table reports the notation for the indices used in the optimization model.

The given data for the problem are the discounted liability cash flows $LCF_{k,c,t}$ for scenario c in the scenario set k at time t and the discounted cash flows $ACF_{k,c,t,a,i}$ for asset i in asset class a for scenario c in scenario set k at time t. These discounted liability and asset cash flows are exogenously given data.

We minimize the sum of the absolute differences between the discounted liability cash flows and the RP cash flows for all years t and all scenarios c in all scenario sets k,

$$min_x \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{t=1}^{T} \left| LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} x_{a,i} ACF_{k,c,t,a,i} \right| \tag{3.2}$$

where the decision variables $x_{a,i}$ are the holdings in each of the $i = 1, 2, \ldots, I_a$ candidate assets in each asset class $a = 1, 2, \ldots, A$.

We employ two standard linear reformulations of this optimization problem in order to obtain numerically tractable linear programs. For this purpose, we first replace each decision variable $x_{a,i}$ by the difference of a nonnegative long position $l_{a,i}$ and a nonnegative short position $s_{a,i}$,

$$x_{a,i} = l_{a,i} - s_{a,i}, \quad l_{a,i}, s_{a,i} \geq 0, \quad \forall\ a, i. \tag{3.3}$$

3.2.2.1 First Linear Model

The first linear reformulation—see Konno and Yamazaki (1991) for an application to equity portfolios—has been previously suggested for RPs; see footnote 1 in Chen and Skoglund (2012). For this reformulation, we introduce a set of auxiliary variables $z_{k,c,t}$ and minimize their sum in the new linear objective function

$$min_{l,s,z} \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{t=1}^{T} z_{k,c,t} \tag{3.4}$$

In addition, we need to introduce constraints relating each auxiliary variable to a term in the original objective function in the L_1 problem (3.2),

$$\begin{aligned} z_{k,c,t} &\geq LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \qquad \forall\ k,c,t & (3.5) \\ -z_{k,c,t} &\leq LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \qquad \forall\ k,c,t & (3.6) \end{aligned}$$

3.2.2.2 Second Linear Model

The second linear reformulation—see Feinstein and Thapa (1993)—has not yet been applied to RPs but appears to be widely used in portfolio optimization. For example, Rudolf et al. (1997) apply it to tracking error minimization. According to Feinstein and Thapa (1993), this approach is highly efficient due to the relatively small number of additional auxiliary constraints. For this reformulation, we introduce for each term in the original L_1 objective function the positive deviation $y^+_{k,c,t}$ and the negative deviation $y^-_{k,c,t}$ as auxiliary variables and minimize their sum in the new linear objective function.

$$min_{l,s,y^+,y^-} \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{t=1}^{T} \left(y^+_{k,c,t} + y^-_{k,c,t} \right) \tag{3.7}$$

Analogously to the first reformulation, we relate the auxiliary variables to the terms in the original objective function via some new constraints,

$$y^+_{k,c,t} - y^-_{k,c,t} = LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \quad \forall\ k,c,t \tag{3.8}$$

$$y^+_{k,c,t}, y^-_{k,c,t} \geq 0 \quad \forall\ k,c,t \tag{3.9}$$

Seemann (2009) argues that an RP may need to have additional "nice" properties for it to be helpful for a thorough analysis of a life insurance portfolio. However, these properties are not reflected in the objective function—that is, the simple minimization of the L_1 (or any other norm) may not lead to an RP that has such favorable properties. For the computation of a truly helpful RP, we therefore introduce additional constraints to the model. In our experience, two sets of constraints are very helpful.

Burmeister and Mausser (2009) explain that an RP should have a relatively small number of assets, because a small RP is relatively quick to price, easy to interpret in relation to the liabilities, and suited to replicating the liability across a wide range of market conditions. In order to limit the number of assets in the optimal solution, we introduce a set of constraints to directly control the size of the complete RP. These constraints have the additional benefit of limiting absolute magnitude of the individual portfolio positions. From a risk perspective, Chen and Skoglund (2012) advise not only to consider the cash flow (mis)match in general but to pay special attention to potential outliers for extreme scenarios. For this purpose, we also impose a set of constraints to ensure a good cash flow match for a broad range of future events.

3.2.3 Additional Constraints

The first set of constraints is established to limit the total number of assets in the portfolio and to avoid so-called offsetting problems. Offsetting refers to the consequences of the well-known multicollinearity problem from regression analysis in the context of RPs. Such an RP exhibits very large long positions in some assets, which are offset by very large short positions in an asset with almost collinear payoffs across the scenarios. While such an RP may offer a slightly improved in-sample fit compared to an RP restricted by bounds, it usually leads to significantly worse out-of-sample fits and it is, therefore, undesirable.

We impose the following linear constraint for each asset class a:

$$nl(\sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t}/C) \geq \sum_{i=1}^{I_a}(l_{a,i} + s_{a,i})nv_a \qquad \forall a, \tag{3.10}$$

with the parameters nl and nv_a. The so-called notional value nv_a for the asset class a is the par value of a financial asset in this class. (For example, the par value of a bond is defined as the amount that the issuer agrees to repay at the maturity date to the bondholder.) These notional values can be different depending on the asset class, but they are the same for all the assets in each class. The notional limit nl determines the largest permitted value for the notional values of an asset class relative to the discounted liability cash flows in the base case. This set of constraints provides bounds on the sums of the absolute values of the short positions $s_{a,i}$ and long positions $l_{a,i}$ of every asset class.

A key feature for a good replicating portfolio is to provide a good match of assets and liabilities for all kinds of events in the future. In our model we try to ensure this match by relating the non-base scenario sets to the base scenario set. We want the discounted asset cash flows in a non-base scenario set to differ from the discounted asset cash flows in the base scenario set in the same way as the discounted liability cash flows in a non-base scenario set differ from the discounted liability cash flows in the base scenario set. Therefore we introduce a set of constraints to limit these scenario set movements between the base scenario set $k = 1$ and the non-base scenario sets $k = 2, \ldots, K$. For each of the non-base scenario sets an upper and a lower bound constraint is established. We impose the following sets of constraints:

$$\begin{aligned} adub \geq \Bigg(&\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i}-s_{a,i}) - \sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i}-s_{a,i}) \\ &- \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \Bigg) / \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.11}$$

$$\begin{aligned} adlb \leq \Bigg(&\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i}-s_{a,i}) - \sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i}-s_{a,i}) \\ &- \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \Bigg) / \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.12}$$

The parameter $adub$ is the upper bound for the scenario set difference and, similarly, $adlb$ is the lower bound for the scenario set difference. The linear inequalities (3.11) and (3.12) must hold for $k = 2, \ldots, K$.

This completes the formulation of our two linear optimization problems. For completion, Appendix 3.A summarizes both complete LPs.

3.2.4 Trimming

Theoretically, our two optimization problems may have multiple optimal solutions. In particular, there may be multiple solutions with different values for $l_{a,i}$ and $s_{a,i}$ for some a, i, but which are otherwise identical. Such a portfolio has unnecessarily offsetting long and short positions in the same asset. Obviously, the more suitable solution has the feature that either $l_{a,i}$ or $s_{a,i}$ is zero. Therefore, we need to calculate the trimmed solution for every feasible solution (z, l, s). In this trimmed solution, $l_{a,i}s_{a,i} = 0$ for $i = 1, \ldots, I$ and $a = 1, \ldots, A$ (see Jacobs et al. (2005)). According to the description of trimming used in Jacobs et al. (2006), we remove the offset from simultaneous long and short positions in our candidate assets in such a way that the smaller of these positions diminishes to zero. This trimmed solution $(z, \bar{l}, \bar{s})$ is defined as follows:

$$\bar{l}_{a,i} = l_{a,i} - \delta_{a,i}, \quad \bar{s}_{a,i} = s_{a,i} - \delta_{a,i}, \quad \text{for all } i = 1, \ldots, I_a, a = 1, \ldots, A, \tag{3.13}$$

where $\delta_{a,i}$ is given by

$$\delta_{a,i} = min\{l_{a,i}, s_{a,i}\} \quad \text{for all } i = 1, \ldots, I_a, a = 1, \ldots, A. \tag{3.14}$$

3.3 Tests

Our two LPs deliver an RP that minimizes the L_1 error (from the liability cash flows) under some additional, desirable constraints. However, a useful replicating portfolio should also have other good properties. Therefore, we now introduce several important tests to check the quality of the optimal replicating portfolio.

3.3.1 Out-of-sample Scenario Set Test

We have already mentioned the importance of the similar behaviors of discounted asset cash flows and discounted liability cash flows for all non-base scenario sets compared to the base scenario set. In the optimization process we can only consider these behaviors in-sample. However, since a pure focus on the in-sample data set leads to the risk of over-fitting for specific data values, we also have to consider scenario set movements for an out-of-sample data set. (Seemann (2009) also recommends the analysis of RPs under new, unknown scenarios.) For this reason, we introduce our first set of tests, the out-of-sample scenario set movement test. The out-of-sample (Oos) movement for all non-base scenario sets $k = 2, \ldots, K$ is denoted by

$$\begin{aligned} Oos_k = \Bigg(\sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF^{os}_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF^{os}_{1,c,t,a,i}(l_{a,i} - s_{a,i}) \\ - \sum_{c=1}^{C} \sum_{t=1}^{T} LCF^{os}_{k,c,t} + \sum_{c=1}^{C} \sum_{t=1}^{T} LCF^{os}_{1,c,t} \Bigg) / \sum_{c=1}^{C} \sum_{t=1}^{T} LCF^{os}_{1,c,t} \end{aligned} \tag{3.15}$$

The right-hand side of (3.15) closely resembles the right-hand side of the constraints (3.11) and (3.12) in our two LPs. However, we substitute the in-sample discounted liability cash flows $LCF_{k,c,t}$ by the out-of-sample discounted liability cash flows $LCF^{os}_{k,c,t}$ and the in-sample discounted asset cash flows $ACF_{k,c,t,a,i}$ by the out-of-sample discounted asset cash flows $ACF^{os}_{k,c,t,a,i}$. As a result, the values Oos_k may violate the bounds *adub* and *adlb* and so may give us an indication of how poorly the optimal RP behaves out-of-sample.

3.3.2 Market Value Movement Test

The scenario sets for the market values include the original scenario sets $k = 1, \ldots, K$ and additional out-of-sample scenario sets. These latter sets are denoted by $m = 1, \ldots, M$. We compare the market value movements of each of the non-base scenario sets $m = 2, \ldots, M$ with the market value movements of the base scenario set $m = 1$. It is very important for a reliable RP that the market value of the RP in a non-base scenario set differs from the market value of the RP in the base scenario set to the same extent as the market value of the liabilities in the non-base scenario set differs from the market value of the liabilities in the base scenario set. The market value movement test for all non-base market value scenario sets $m = 2, \ldots, M$ is denoted by

$$MV_m = \left(\sum_{a=1}^{A} \sum_{i=1}^{I_a} MVA_{m,a,i}(l_{a,i} - s_{a,i}) - MVA_{1,a,i}(l_{a,i} - s_{a,i}) - (MVL_m - MVL_1) \right) / MVL_1 \quad , \tag{3.16}$$

where $MVA_{1,a,i}$ and $MVA_{m,a,i}$ are the market values for asset class a and candidate asset i in the base scenario and for all non-base scenarios $m = 1, \ldots, M$. MVL_1 and MVL_m are the values for the market value liabilities in the base scenario and for all non-base scenario sets $m = 1, \ldots, M$.

3.3.3 The Coefficient of Determination

According to Seemann (2009) and Devineau and Chauvigny (2011) the coefficient of determination, R^2, is commonly used as an indicator for assessing the quality of an RP. The R^2 can also be found in Daul and Gutiérrez Vidal (2009) and Burmeister and Mausser (2009) as a method for testing the quality of an RP. And so, in our last test, we calculate the in-sample and the out-of-sample R^2 values. In addition we calculate the annual coefficient of determination for every year t separately. The in-sample R^2 is

$$R^2_{is} = \frac{\sigma(LCF, ACF)^2}{\sigma^2(LCF)\sigma^2(ACF)} \quad , \tag{3.17}$$

where the point estimates for the variance of the liabilities, the variance of the RP, and the covariance between the RP and the liabilities are calculated as follows:

$$\sigma^2(LCF) = \frac{\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T}(LCF_{k,c,t} - \overline{LCF})^2}{(K * C * T - 1)} \tag{3.18}$$

$$\sigma^2(ACF) = \frac{\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a}(ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \overline{ACF})^2}{(K * C * T - 1)} \tag{3.19}$$

$$\sigma(LCF, ACF) = \frac{\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T}(LCF_{k,c,t} - \overline{LCF})(\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \overline{ACF})}{(K * C * T - 1)} \tag{3.20}$$

with

$$\overline{LCF} = \sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T} LCF^{os}_{k,c,t}/(K * C * T) \tag{3.21}$$

and

$$\overline{ACF} = \sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF^{os}_{c,t,i}(l_{a,i} - s_{a,i})/(K * C * T) \tag{3.22}$$

The out-of-sample R^2 is calculated in the same way as the in-sample R^2; we merely replace the in-sample discounted liability and asset cash flows by the out-of-sample discounted liability and asset cash flows

$$R^2_{os} = \frac{\sigma(LCF^{os}, ACF^{os})^2}{\sigma^2(LCF^{os})\sigma^2(ACF^{os})} \tag{3.23}$$

We also calculate the annual coefficient of determination using the out-of-sample data for the particular year t. The annual R^2 is an appropriate feature for analyzing the fit per year. Years with a low R^2 have a negative impact on the performance of the RP. A way of improving the coefficient of determination for these years could be to introduce new candidate assets with cash flows in these particular years. Including these assets and rerunning the optimization may result in a better annual R^2 and a better RP.

[5]I thank Zurich for providing us with the data sets. I am especially grateful to Regine Scheder and Lucio Fernandez Arjona for professional advice and helpful discussions regarding the topic. For reasons of confidentiality, all data sets are slightly modified versions of real-life data sets.

3.4 Data Sets and Specific Problem Instance

We solve our models developed in Section 3.2 using a data set provided to us by Zurich.[5] Furthermore we have two additional data sets provided by Zurich for the Oos scenario set test and the market value movement test introduced in Section 3.3, respectively.

3.4.1 Data Sets: General Description

The first data set contains the in-sample data and is the basis for the RP optimization. This data set contains discounted liability cash flows for all C scenarios in each of the K scenario sets across the entire time horizon T. Furthermore, this data set also provides discounted cash flows for all C scenarios in each of the K scenario sets across the time horizon T for I_a preselected candidate assets in each asset class $a = 1, 2, \ldots, A$. (The selection of the particular candidate assets for the optimization is based on an interaction between local business units of Zurich and the RP team at group level. As a starting point for an appropriate set of candidate assets, final compositions of previous RPs are typically used.)

The data for the scenario set $k = 1$ represents the base scenario. The other scenario sets, $k = 2, \ldots, K$, consider changes in particular market factors that could influence both the asset and liability cash flows. These scenario sets are generated using an economic scenario generator (ESG) model, which is—according to Varnell (2011)—an essential element for ensuring a market-consistent valuation in the life insurance business. The ESG models at Zurich (as do all ESG models in general) rely on Monte Carlo simulations. After the scenario set creation, the discounted cash flows of the liabilities are generated based on the scenario sets. This is done according to the principles of market-consistent embedded value (MCEV).[6] In addition, the ESG models also provide discounted cash flows for the preselected candidate assets for the same economic scenarios as used for the liabilities.

The second data set contains the out-of-sample data. This data for the Oos scenario set movement test is generated in the same way as the set for the optimization. Finally, the data set for the market value test provides Monte Carlo prices for each candidate asset in each asset class for all market value scenario sets $m = 1, 2, \ldots, M$.

3.4.2 Specific Problem Instance

We next describe the specific data sets. The in-sample data set comprises $K = 13$ scenario sets with $C = 100$ scenarios for each year over a time horizon of $T = 40$ years. The set of candidate assets consists of $A = 5$ different asset classes with a total of 919 assets. The out-of-sample data set has the same dimensions as the in-sample data set. Finally, the out-of-sample market value data set contains $M = 15$ market value scenario sets. These scenario sets are average values based on several simulations and do not contain any actual scenarios. The first 13 scenario sets consider the same

[6] So-called embedded values are used to report the value of a life insurance company. MCEVs increase the consistency and the disclosure of embedded value reporting in the European insurance industry. For more information see CFO Forum (2012).

economic conditions as the two cash flow data sets; the last two scenario sets contain scenarios for further economic shocks.

3.4.2.1 Scenario Sets

Table 3.2 describes the economic conditions characterizing the $K = 13$ scenario sets. The first of the

Table 3.2: List of Scenario Sets.

Scenario set	Notation	Feature
1	$k = m = 1$	Base
2	$k = m = 2$	Yield curve up small
3	$k = m = 3$	Yield curve down small
4	$k = m = 4$	Yield curve up big
5	$k = m = 5$	Yield curve down big
6	$k = m = 6$	Equity index down big
7	$k = m = 7$	Property index down big
8	$k = m = 8$	Equity index down small
9	$k = m = 9$	Property index down small
10	$k = m = 10$	Interest rate volatility up
11	$k = m = 11$	Interest rate volatility down
12	$k = m = 12$	Equity index volatility up
13	$k = m = 13$	Equity index volatility down
14	$m = 14$	Shock 1
15	$m = 15$	Shock 2

This table reports scenario sets for the optimization and the tests.

$K = 13$ scenario sets is the base scenario. This scenario set is in line with current market prices. In addition, four scenario sets take a shift of the yield curve into account. Two of these simulate a small and a large up-shift of the yield curve, respectively. The other two scenario sets simulate a small and a large down-shift of the yield curve, respectively. Furthermore, two scenario sets simulate a small and a large down-shift of the main equity index, respectively. (The main equity index is the most important equity index in the currency of the life insurance contract. For example, for a life insurance contract in Swiss francs the main equity index is the Swiss Market Index (SMI) of the 20 largest companies in Switzerland.) The next two scenario sets simulate a small and a large down-shift of the main property index, respectively. Moreover, two scenario sets reflect an up- and a down-shift in the interest rate volatility, respectively and two more scenario sets simulate an up- and a down-shift in the main equity index volatility, respectively. Finally, there are two out-of-sample scenario sets for the market value test. These more extreme scenario sets simulate two different kinds of economic shock.

3.4.2.2 Asset Universe

Table 3.3 describes the set of candidate assets. We consider 919 candidate assets separated into $A = 6$ asset classes suggested to us by Zurich. All of these assets can be "priced" under various

Table 3.3: Available Asset Universe.

Asset class	Notation	Available assets
Zero coupon bonds	$a = 1$	$I_1 = 40$
Equity indices	$a = 2$	$I_2 = 40$
Property indices	$a = 3$	$I_3 = 40$
Cash indices	$a = 4$	$I_4 = 40$
Interest rate swap payer assets	$a = 5$	$I_5 = 39$
Pay fixed swaptions	$a = 6$	$I_6 = 720$

This table reports the available asset universe for the optimization.

market conditions, even though some of them are not traded on deep and liquid financial markets. The inclusion of such hypothetical assets increases the flexibility and the quality of the replication results. We only provide a brief description of these asset classes; for more detailed information see Fabozzi (2008).

Zero coupon bonds do not make periodic coupon (interest) payments. The holder of such a bond generates interest income by buying it below its par value. The face value of the bond is paid at the time of its maturity and the aggregated interest realized for the holder is the par value minus the purchase price. Investments in equity indices and property indices are assumed to be sold at a specific, predetermined point in time. The cash flow, which only happens at this preset point in time, is based on the value of the underlying index. The equity index is usually the most important index in the currency of the liabilities. The property index can be regarded as a synthetic real estate index in the currency of the liabilities. Cash indices are cash accounts, which are cumulating the annual interest until maturity and then pay out. Interest rate swap payer assets pay a fixed interest payment to another party at designated dates for the life of a specific contract. They are an efficient hedge against short-term interest rate increases (see Fabozzi and Buetow (2008a)). Pay fixed swaptions are options on interest rate swaps. This swaption type grants its holder the right to enter into an interest rate swap in which he or she pays a fixed rate and receives a floating rate (see Fabozzi and Buetow (2008b)).

3.4.2.3 Model and Test Parameters

Table 3.4 provides an overview of the parameters for the optimization model, the out-of-sample test, and the market value test. Most of the parameters for our optimization model were provided to us. In total we have 52,000 discounted liability cash flows (one for each scenario in every scenario set in

Table 3.4: Parameter Overview.

Optimization Model—Provided Parameters	
$LCF_{k,c,t}$	52,000 discounted liability cash flows
$ACF_{k,c,t,a,i}$	52,000 discounted asset cash flows for each of the 919 assets
$nv_1 = 100$	notional value for asset class 1
$nv_2 = 6822.44$	notional value for asset class 2
$nv_3 = 100$	notional value for asset class 3
$nv_4 = 100$	notional value for asset class 4
$nv_5 = 100$	notional value for asset class 5
$nv_6 = 100$	notional value for asset class 6
Optimization Model—Determined Parameters	
$nl = 3$	notional limit for asset class 6
adub	17 allowed differences upper bounds between 0.008 and 0.001
adlb	17 allowed differences lower bounds between -0.008 and -0.001
Tests—Out-of-sample	
$LCF^{os}_{k,c,t}$	52,000 discounted out-of-sample liability cash flows
$ACF^{os}_{k,c,t,a,i}$	52,000 discounted out-of-sample asset cash flows for each of the 919 assets
Tests—Market Value	
MVL_m	15 market value liabilities
$MVA_{m,a,i}$	15 market values for each of the 919 assets

This table provides a parameter overview for the optimization model, the out-of-sample test, and the market value test.

each year) denoted by $LCF_{k,c,t}$. We also have 52,000 discounted cash flows (one for each scenario in every scenario set in each year) for each of the 919 available assets. They are denoted by $ACF_{k,c,t,a,i}$. The notional value for asset classes 1, 3, 4, 5, and 6 is 100 while the notional value for asset class 2 is 6,822.44. (This notional value is based on the value of a certain index on a reference date.)

In addition, we have to determine several parameters for our constraints (3.10), (3.11), and (3.12). As mentioned in Section 3.2, we want to avoid offsetting problems and limit the total number of assets in the portfolio by introducing a notional limit nl. The notional limit should be in force for all asset classes. However, empirical observations showed that the available assets in the asset classes $a = 1, \ldots, 5$ did not reach the notional limit even without the additional constraint. Therefore we did not use notional limit constraints for these asset classes. (The absence of the notional limit constraints for these asset classes does not have an effect on the numerical results. It only enhances the computational performance and introducing such constraints would be easily possible.) Since we have a rather large number of pay fixed swaptions in our set of candidate assets, we limit the notionals in asset class $a = 6$ according to (3.10). We use a notional limit of $nl = 3$, which is in line with the value usually used in the life insurance industry.

We also have to determine values for the bounds $adub$ and $adlb$ in (3.11) and (3.12). We start with the rather large values $adub = 0.008$ and $adlb = -0.008$, so we tolerate a scenario set difference between -0.8% and 0.8%. For these values the objective value only changes slightly compared to the case of an unconstrained optimization. Subsequently, we synchronously reduce $adub$ by 0.0005 and increase $adlb$ by 0.0005 until we reach $adub = 0.001$ and $adlb = -0.001$. Since we find the best value for the market value movement test at $adub = 0.002$ and $adlb = -0.002$, we run additional calculations for $adub = 0.00175$ and $adlb = -0.00175$ and $adub = 0.00225$ and $adlb = -0.00225$.

We also have parameters for the out-of-sample scenario set movement test and the market value test. These parameters were also provided by Zurich. For the Oos scenario set movement test we have 52,000 discounted out-of-sample liability cash flows denoted by $LCF^{os}_{k,c,t}$ and 52,000 discounted out-of-sample cash flows for each of the 919 assets denoted by $ACF^{os}_{k,c,t,a,i}$. In addition, we have parameters for the market value test; these are the market values MVL_m for each of the 15 market value scenario sets and the market values $MVA_{m,a,i}$ for each of the 919 candidate assets for the 15 different scenario sets.

We implement our optimization models using GAMS. Firstly, we prepare the provided input data using Microsoft Excel. Secondly, we formulate and solve our optimization problem in GAMS and export our results to Excel. And finally, we use Excel to test the quality of our results. Appendix 3.B describes the details of the data handling and the implementation.

3.5 Results

We first compare the test results for the optimal RPs as a function of the absolute allowed scenario set difference $adub = -adlb$ (AAD). Next we depict in detail the optimal RPs for three particular values of the AAD: 0.175%, 0.25%, and 0.45%. Finally, we compare the computing performance of the two different linear programming models introduced in Section 3.2.

3.5.1 Comparison of Different RPs

Table 3.5 provides detailed test results for the optimal RPs as a function of the AAD. The first row of results, marked with "UC" for unconstrained, reports results for the minimization of the L_1 error *without* the additional constraints (3.10)–(3.12). Obviously, the optimal L_1 error for this problem, $L_1^{UC} = 1.258 \times 10^9$, serves as a lower bound on the L_1 errors of the constrained problems. The remaining seventeen rows of Table 3.5 report results for constrained optimization problems. The values of the bounds AAD decrease from 0.008 to 0.001. The two columns under the heading RL_1 report the relative deviation of the in-sample and out-of-sample L_1 errors, respectively, from the in-sample error L_1^{UC} for the unconstrained problem. The two columns under the heading MV Mov report the average and worst-case market value movements while the two columns under the heading Oos SS Mov report the average and worst-case out-of-sample scenario set movements. The in-sample and out-of-sample R^2 values can be found under the heading R^2. (As we would expect, both linear reformulations lead to the same results. Therefore, we do not need to differentiate between the two models in our discussion of the results in Sections 3.5.1 and 3.5.2. However, in Section 3.5.3 we observe that the type of reformulation matters for running times.)

Table 3.5: Results for different RPs.

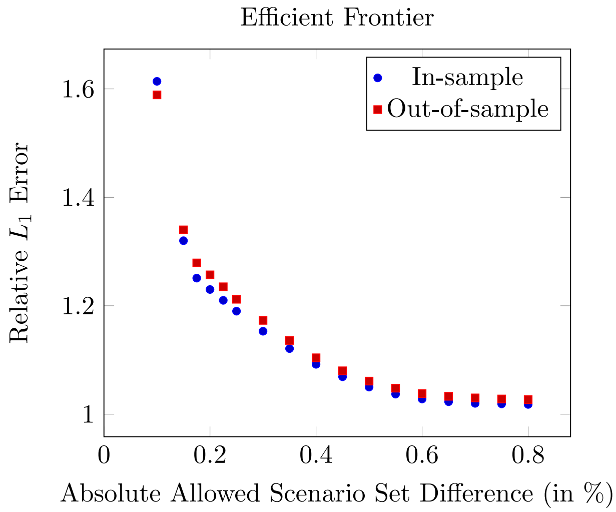
	RL_1		MV Mov		Oos SS Mov		R^2	
AAD	Is	Oos	Av	WC	Av	WC	Is	Oos
UC	1	1.037	0.676%	2.500%	0.568%	1.417%	98.917%	98.823%
0.8%	1.018	1.027	0.616%	2.089%	0.555%	1.434%	98.896%	98.848%
0.75%	1.019	1.028	0.600%	2.035%	0.545%	1.397%	98.892%	98.847%
0.7%	1.020	1.030	0.589%	1.996%	0.531%	1.354%	98.887%	98.840%
0.65%	1.023	1.033	0.572%	1.941%	0.514%	1.277%	98.884%	98.831%
0.6%	1.028	1.038	0.556%	1.914%	0.496%	1.143%	98.875%	98.817%
0.55%	1.037	1.048	0.538%	1.890%	0.470%	1.031%	98.857%	98.773%
0.5%	1.050	1.061	0.504%	1.811%	0.444%	0.911%	98.828%	98.729%
0.45%	1.069	1.080	0.463%	1.741%	0.412%	0.718%	98.790%	98.664%
0.4%	1.092	1.104	0.421%	1.665%	0.362%	0.734%	98.729%	98.541%
0.35%	1.121	1.136	0.382%	1.597%	0.318%	0.750%	98.672%	98.450%
0.3%	1.153	1.173	0.348%	1.529%	0.283%	0.754%	98.589%	98.316%
0.25%	1.190	1.212	0.325%	1.467%	0.274%	0.768%	98.460%	98.164%
0.225%	1.210	1.235	0.327%	1.436%	0.293%	0.762%	98.370%	98.044%
0.2%	1.230	1.257	0.322%	1.401%	0.299%	0.771%	98.265%	97.897%
0.175%	1.251	1.279	0.319%	1.367%	0.308%	0.775%	98.141%	97.804%
0.15%	1.320	1.340	0.332%	1.323%	0.372%	0.847%	97.502%	97.370%
0.1%	1.614	1.589	0.382%	1.281%	0.517%	1.167%	94.583%	95.651%

This table reports results for RL_1 errors, market value movements, Oos scenario set movements, and the in-sample and out-of-sample R^2 for the unconstrained RP, and seventeen RPs with different values in the AAD. AAD = absolute allowed scenario set difference, Mov = movement, SS = scenario set, Oos = out-of-sample, Is = in-sample, Av = average, WC = worst case, Corr = total correlation, UC = unconstrained.

3.5.1.1 Effects of the Absolute Allowed Scenario Set Difference

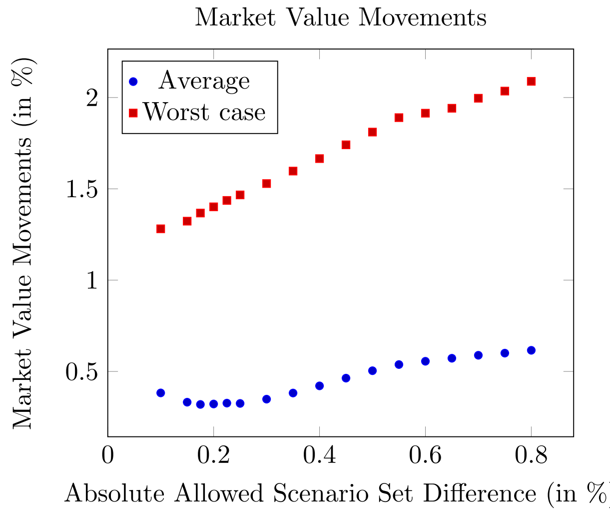
Figure 3.1 displays the results reported in the two columns under the heading RL_1 from Table 3.5; it shows the effect of the AAD on the RL_1 values. Perhaps somewhat surprisingly, the in-sample and out-of-sample RL_1 values are similar for all values of *adub* and *adlb*. As expected, the RL_1 error values slowly increase when we tighten the constraints (3.11) and (3.12) by reducing the AAD from 0.8% to 0.15%. However, there is a large increase in the RL_1 values between an AAD of 0.15% and

Figure 3.1: Relative Errors.



Effect of the AAD on the relative in-sample and relative out-of-sample L_1 errors.

Figure 3.2: Market Value Movements.



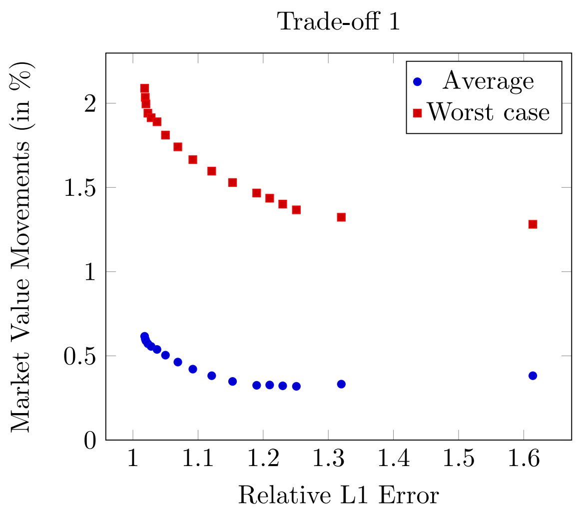
Effect of the AAD on the average and worst-case scenario set market value movements.

0.1%. This effect shows us the high price of the reduction of the AAD after a certain level.

Figure 3.2 displays the results reported in the two columns under the heading "MV Mov" from Table 3.5; it shows the effect of the AAD on the market value movements. The (red) squares represent the largest movements across the 14 non-base scenario sets. For all values of the AAD, the scenario set simulating shock 1 ($m = 14$) has the highest market value movement. The maximal market value movement decreases monotonically from 2.089% at an AAD of 0.8% to 1.281% at an AAD of 0.1%. The (blue) dots show the average market value movements for the 14 non-base scenarios depending on the AAD. This average market value movement decreases from 0.616% at an AAD of 0.8% to a minimum of 0.319% at an AAD of 0.175%. A further reduction of the AAD to 0.1% leads to an increase of the average market value movement to 0.383%. Considering the effects of the AAD we see a trade-off between small RL_1 errors and small market value movements. This trade-off is shown in Figure 3.3. An increase of the RL_1 error always leads to a decrease of the market value movement of the worst scenario set. However, the decrease from 1.323% to 1.281% comes with an RL_1 error increase from 1.320 to 1.614. Small values for the average market value movement have the price of an increase of the RL_1 error. The smallest average market value movement we find is 0.319% and it comes with an RL_1 error of 1.251. Further increases of the RL_1 error do not have the benefit of a decrease of the average market value movement. So the two blue dots on the right side of Figure 3.3 are inefficient.

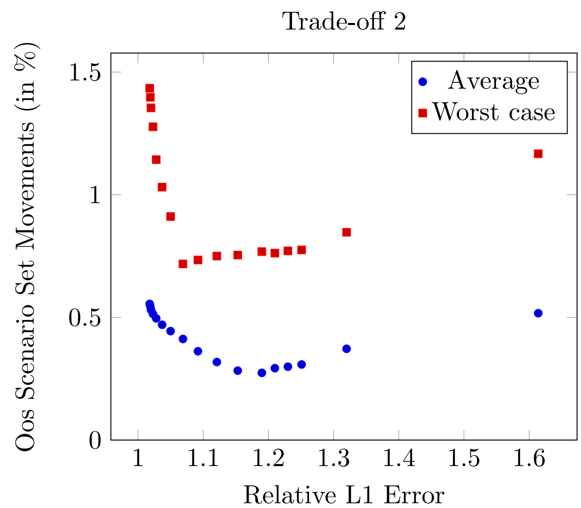
In a next step we examine the Oos scenario set movements of several RPs. Figure 3.4 shows a trade-off between a small RL_1 error and a small Oos scenario set movement. The smallest RL_1 error, 1.018, comes with the highest values for the average of the Oos scenario set movement and the highest value for the worst of the twelve non-base scenario sets. However the smallest Oos scenario set movement value for the worst scenario set is already reached at a relatively small RL_1 error

Figure 3.3: Market Value Movements vs. Relative Errors.



Trade-off between market value movements and relative L_1 in-sample errors.

Figure 3.4: Oos Scenario Set Movements vs. Relative Errors.



Trade-off between Oos scenario set movements and relative L_1 in-sample errors.

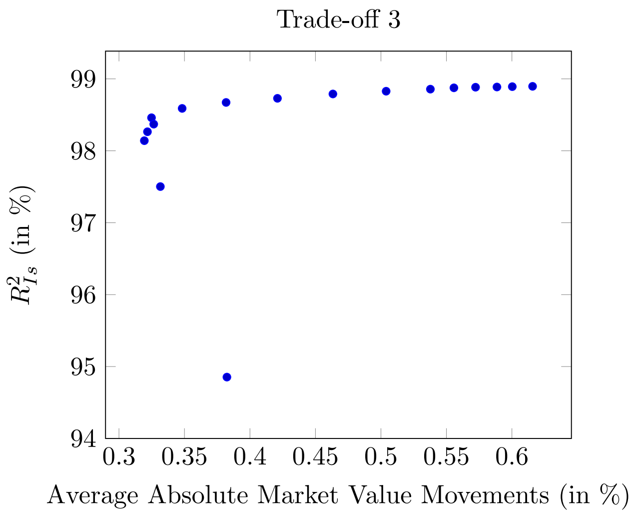
of 1.069. A further increase of the RL_1 error has a negative effect on the Oos movement of the worst scenario set. A similar, but less extreme behavior can be seen for the average Oos scenario set movement. The smallest value for the average Oos scenario set movement comes with an RL_1 error of 1.121. A further increase of the RL_1 error leads to a higher average Oos scenario set movement.

Analyzing the trade-offs shown in Figure 3.3 and Figure 3.4, we see the danger of very small AADs. Reducing the AADs under 0.175% has a small negative effect on the average market value movement and a large negative effect on the RL_1 error. For the Oos scenario set movement AADs should even not be smaller than 0.25%.

3.5.1.2 Total Coefficient of Determination

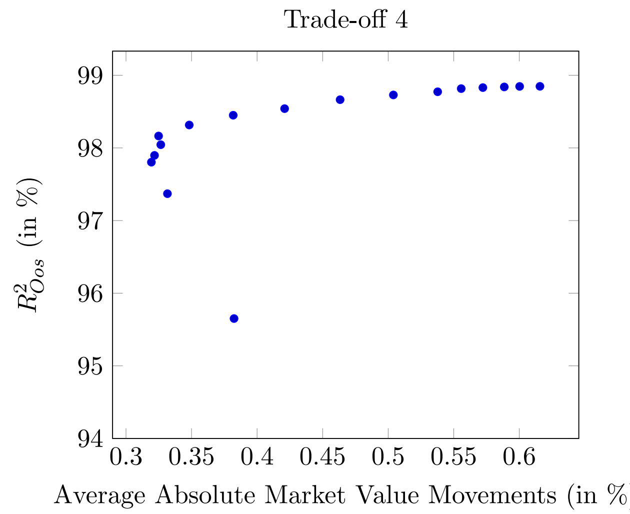
Another indicator for assessing the quality of a good RP is a high coefficient of determination over the total 40 years in-sample and out-of-sample. In Figure 3.5 we see the trade-off between a high total R^2_{Is} and a small average market value movement. At an AAD of 0.8% the total R^2_{Is} has a good value of 98.896% while the average absolute market value movement is 0.616. A reduction of the AAD leads to a decrease in the total R^2_{Is} and the average absolute market value movements until we have a still acceptable R^2_{Is} of 98.141% and the smallest average absolute market value movement of 0.319 at an AAD of 0.175%. Therefore we are facing a trade-off between a high R^2_{Is} and a small average absolute market value movement for 14 out of 17 RPs. However, we also have three RPs that are inefficient considering only average market value movements and the total R^2_{Is}: the RP at an AAD of 0.225% has a higher average absolute scenario set movement and a smaller R^2_{Is} than the RP at an AAD of 0.25%, and the RPs at an AAD of 0.15% and 0.1% have higher average absolute scenario set movements and smaller R^2_{Is} values than the RP at an AAD of 0.175%.

Figure 3.5: In-sample R^2 vs. Market Value Movements.



Trade-off between the in-sample R^2 and the average absolute market value movements.

Figure 3.6: Out-of-sample R^2 vs. Market Value Movements.



Trade-off between the out-of-sample R^2 and the average absolute market value movements.

The trade-off between the out-of-sample coefficient of determination and the market value movements shown in Figure 3.6 is very similar to the trade-off between the in-sample coefficient of determination and the market value movements. At an AAD of 0.8% we have an R^2_{Oos} of 98.848% and an average market value movement of 0.616%. A decrease of the AAD leads to a slow decrease of the R^2_{Oos} and a decrease of the average market value movement. At an AAD of 0.175% we have a R^2_{Oos} of 97.804% and an average market value movement of 0.319%. The RPs at AADs of 0.225%, 0.15%, and 0.1% are inefficient considering only average absolute market value movements and R^2_{Oos} values.

3.5.2 Detailed Results for Three Replicating Portfolios

After analyzing the four trade-offs of the previous subsection we take a closer look at three different RPs. We decided to evaluate the RPs we found using an AAD of 0.175%, 0.25%, and 0.45%, respectively.

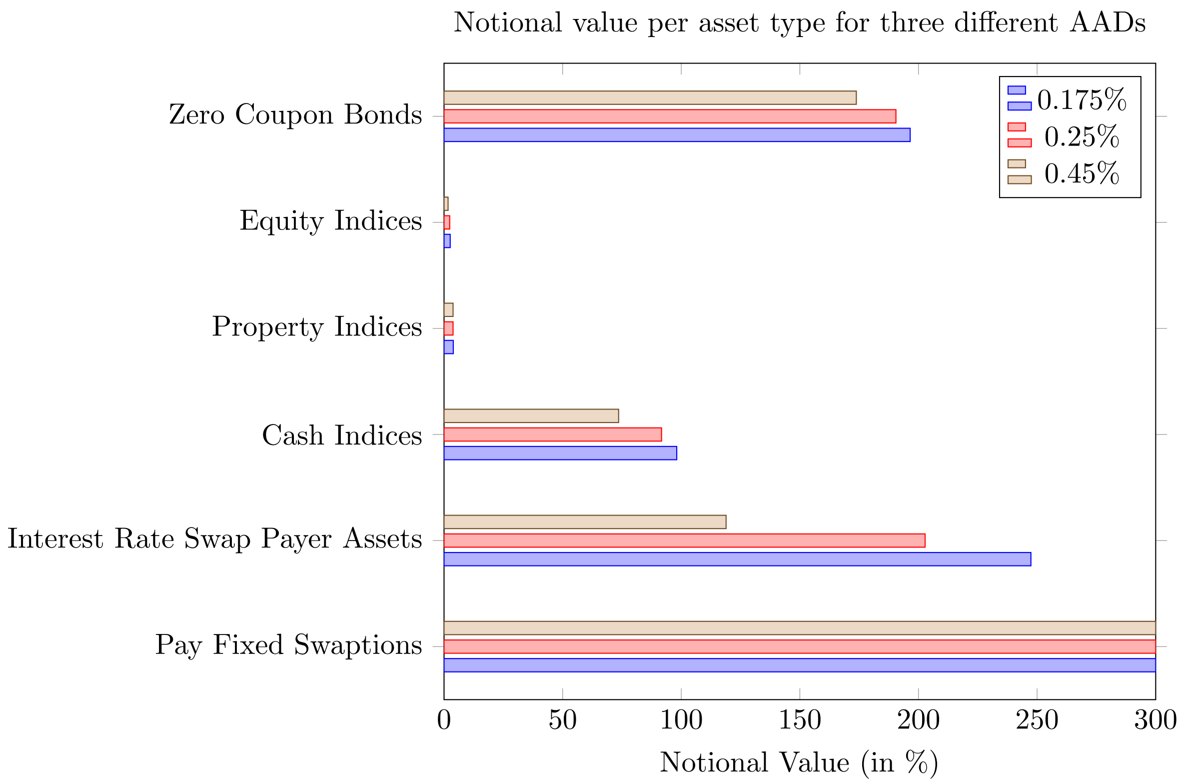
We chose the RP at an AAD of 0.175% because this portfolio has the smallest average market value movement (0.319%) of all RPs. It also has very good values for the worst case market value movement scenario set (1.367%), the average Oos scenario set movement (0.308%), and the worst-case Oos scenario set movement (0.775%). Furthermore the R^2_{Is} (98.141%), the R^2_{Oos} (97.804%), and the RL_1 (1.251) of this RP are still acceptable.

We chose the RP at an AAD of 0.25% because this portfolio has the smallest average Oos scenario set movement (0.274%). The values for the worst-case market value movement scenario set (1.467%), the average market value movement (0.325%), and the worst-case Oos scenario set movement (0.768%) are also very good for this RP. In addition it has a higher R^2_{Is} (98.460%) and

R^2_{Oos} (98.164%) and a lower RL_1 (1.190) than the RP at an AAD of 0.175%.

We chose the RP at an AAD of 0.45% because this portfolio has the smallest worst-case Oos scenario set movement (0.718%). Its values for the worst-case market value movement scenario set (1.741%), the average market value movement (0.463%), and the average Oos scenario set movement (0.412%) are worse than for the other two RPs but still acceptable. However, it has the best R^2_{Is} (98.790%), R^2_{Oos} (98.664%), and RL_1 (1.069) values of our three chosen RPs.

Figure 3.7: Notional Values (in %).



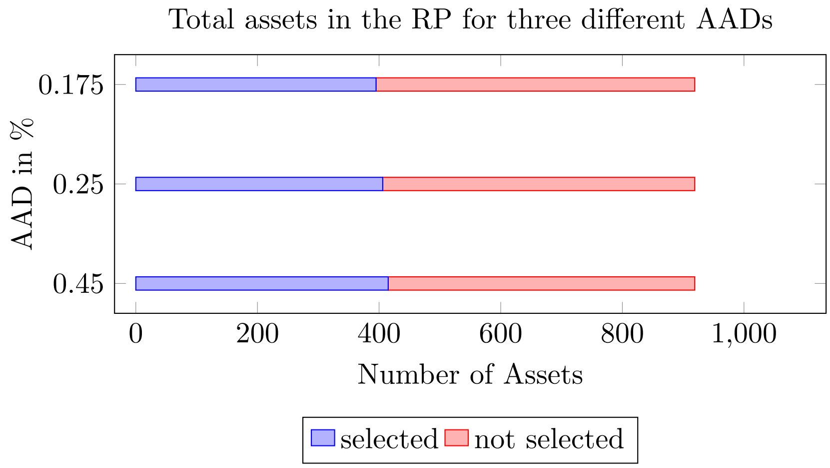
Notional values in % of the market value of the liabilities in the base scenario set for each of the six asset classes using an AAD of 0.175%, 0.25%, and 0.45%, respectively.

3.5.2.1 Composition

First we have a look at the composition of our three RPs. Their notional values for the six asset classes in our set of candidate assets as a percentage of the market value of the liabilities in the base scenario set are shown in Figure 3.7. In our model the total notional values per asset class are the sum of the absolute values of the short and long holdings in an asset class multiplied by the notional value of the asset class.

In our optimization process we tried to limit the size of the portfolio to avoid offsetting problems. In (3.10) we limited the maximum notional value per asset class to 300% of the market value of the liabilities in the base scenario. In practice however we only used this constraint for the pay fixed

Figure 3.8: Number of Selected Assets.



Number of selected and not selected assets in the RP in total using an AAD of 0.175%, 0.25%, and 0.45%, respectively.

swaptions. For each of the three RPs none of the six asset classes exceed the 300% limit. However for all the three RPs the pay fixed swaptions exactly reach the 300% bound.

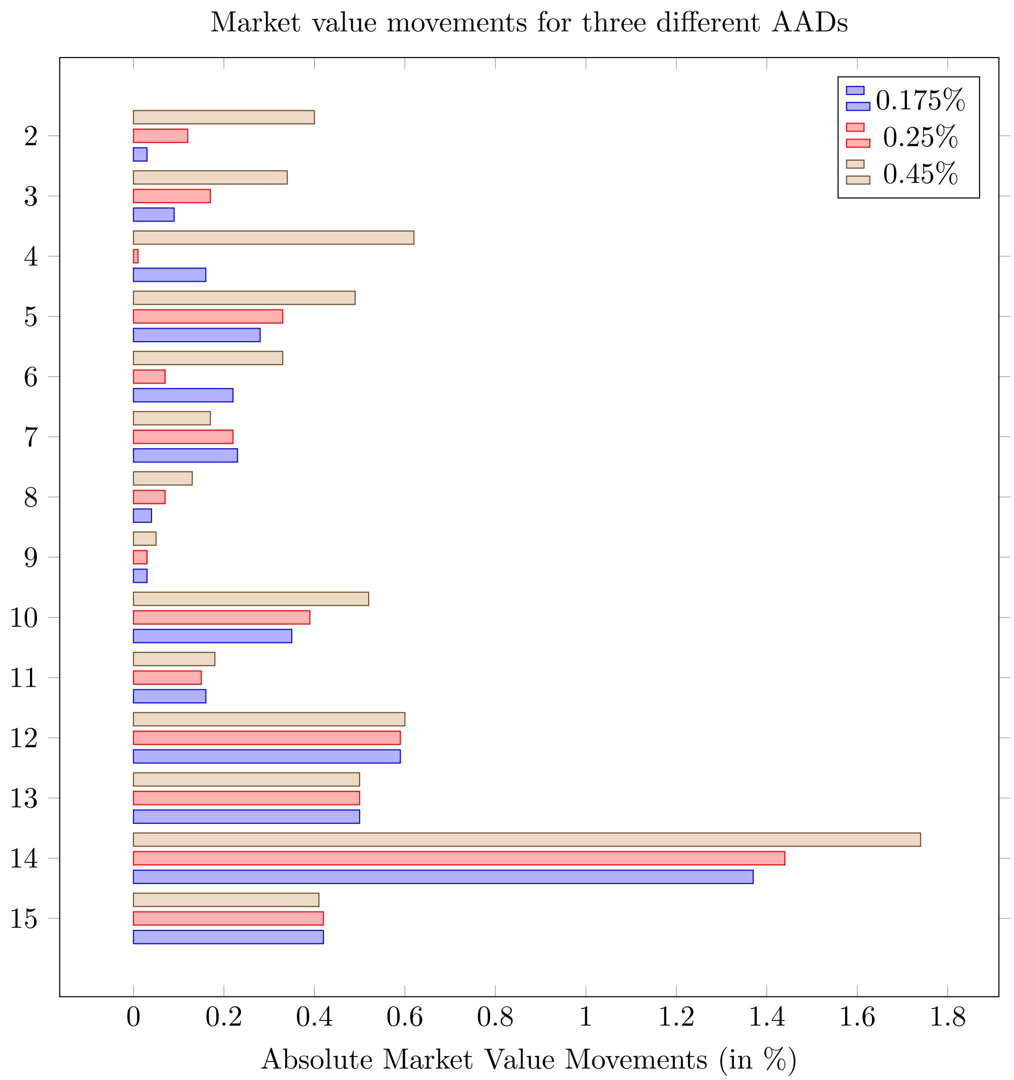
For all asset classes except the pay fixed swaptions the notional values decrease when the AAD increases. For the zero coupon bonds, the equity indices, property indices, and the cash indices the differences in notional values between the three RPs are relatively small. However, the notional values of the interest rate swap payer assets decrease significantly with an increase of the AAD.

According to Burmeister and Mausser (2009) small RPs are generally better than RPs that include all candidate assets because they are easier to price and interpret in relation to the liability. Therefore the creation of an RP with a relatively small number of assets was also one of our optimization goals. Figure 3.8 shows that this goal is accomplished for each of the three RPs. Each include less than half of the candidate assets. However, with an increase of the AAD the number of the candidate assets in the RP increases slightly. The RP with an AAD of 0.175% selects only 395 out of 919 candidate assets, the RP with an AAD of 0.25% selects 406 out of 919 candidate assets, and the RP with an AAD of 0.45% selects 415 out of 919 candidate assets. For each of the three RPs we select all assets in the zero coupon bonds, the equity and property index, the cash indices, and the interest rate swap payer assets asset classes. However, in the pay fixed swaptions asset class we only select 196 out of 720 assets for the RP with an AAD of 0.175%, 207 out of 720 assets for the RP with an AAD of 0.25%, and 216 out of 720 assets for the RP with an AAD of 0.45%.

3.5.2.2 Market Value and Oos Scenario Set Movements

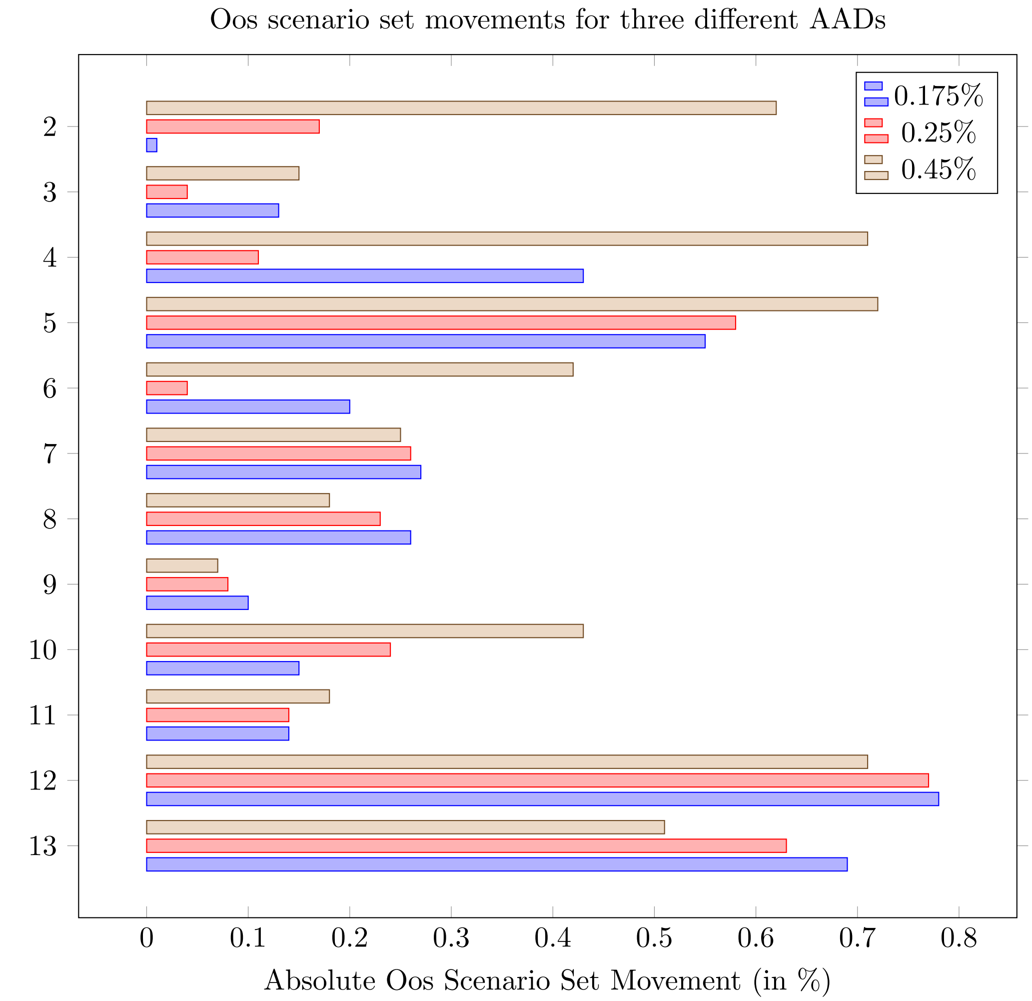
In Figure 3.9 we see the absolute market value movements for the 14 non-base scenario sets in our three RPs. For each of the three RPs the only scenario set with a market value movement larger than one is scenario set 14. The worse performance compared to the other scenario sets can be

Figure 3.9: Absolute Market Value Scenario Set Movements.



Absolute market value scenario set movements for the 14 non-base scenarios using AADs of 0.175%, 0.25%, and 0.45%, respectively.

Figure 3.10: Absolute Oos Scenario Set Movements.



Absolute Oos scenario set movements for the twelve non-base scenario sets using AADs of 0.175%, 0.25%, and 0.45%, respectively.

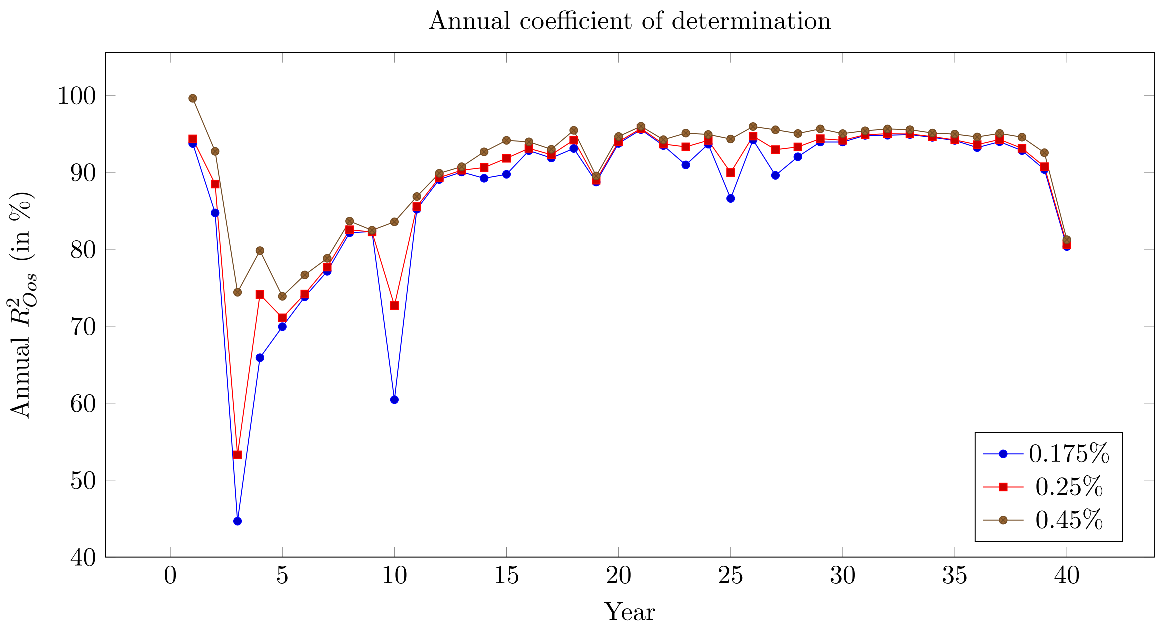
explained by the fact this scenario set is an out-of-sample scenario set. However the absolute market value movements of 1.367% for the RP with an AAD of 0.175%, 1.467% for the RP with an AAD of 0.25%, and 1.741% for the RP with an AAD of 0.45% are still good values for such an extreme scenario set that was not included in the optimization process. The absolute value of the movements of the 13 other non-base scenario sets are all smaller than 0.7% for each of the three RPs. This means the assets and the liabilities in all those scenario sets will most likely behave very similarly to the base scenario. However, on average the RP with an AAD of 0.175% performs better than the RP with an AAD of 0.25%, which in turn performs better on average than the RP with an AAD of 0.45%.

We also obtain very good values in the Oos scenario set movements test for our three RPs. Figure 3.10 shows the absolute values of the Oos movements of the twelve non-base scenario sets for each of the three RPs. For every scenario set and each RP the absolute value of the Oos movement is smaller than 0.8%. These small Oos scenario set movements secure a high probability of an RP performing well in a big range of different future scenarios. However, the RP with an AAD of 0.25% on average performs better than the RP with an AAD of 0.175% and much better than the RP with an AAD of 0.45%.

3.5.2.3 Annual Coefficient of Determination

In Subsection 3.5.1 we examined the total in-sample and out-of-sample coefficients of determination of our RPs. Thereby we found good total R^2_{Is} and R^2_{Oos} values for all our three RPs. For a more detailed analysis we now examine the annual out-of-sample coefficients of determination of the three RPs. Figure 3.11 shows the annual R^2_{Oos} for each of the 40 years in the model for the three RPs.

Figure 3.11: Annual R^2 Values.



Annual out-of-sample R^2_{Oos} using AADs of 0.175%, 0.25%, and 0.45%, respectively.

We see that for all the three RPs most of the annual R^2_{Oos} are significantly worse than their total R^2_{Oos}. We can also see that the RP with an AAD of 0.45% has a higher R^2_{Oos} in every year than the two other RPs. For most of the years it is only slightly higher, however there is a sizable difference in the annual R^2_{Oos} in year three and year ten between the RP with an AAD of 0.45% and the RPs with AADs of 0.175% and 0.25%. Furthermore, for the RP with an AAD of 0.45% all annual R^2_{Oos} are higher than 73%. The RP with an AAD of 0.175% in contrast has three annual R^2_{Oos} smaller than 70% and an extremely small annual R^2_{Oos} of 44.667% in year three. The RP with an AAD of 0.25% only has one annual R^2_{Oos} below 70%, however the annual R^2_{Oos} of 53.281% in year three is also troubling for this RP. The R^2_{Oos} in year three (and the other years with a very small R^2_{Oos}) could be improved by including new assets with cash flows in these particular years and a performing a rerun of the optimization model.

3.5.2.4 Selecting an RP

In Subsection 3.5.1 we presented 17 RPs that we found using different AADs. We then showed their trade-off between good RL_1, R^2_{Is}, and R^2_{Oos} values on one side and good average and worst-case market value and Oos scenario set movements on the other side. We showed that the RPs with AADs of 0.225%, 0.15%, and 0.1% are inefficient considering only R^2_{Is}, R^2_{Oos}, and average market value movements. Furthermore, some RPs showed considerably worse performance in at least one quality criterion than did the others. However none of the RPs was dominated by another RP in every single quality criterion.

In Subsection 3.5.2 we presented a more detailed analysis of three RPs. We showed that their composition and notional values are relatively similar. The three RPs also did not differ extremely in their market value and Oos scenario set movements. We showed, however, that there is an unsatisfactory annual R^2_{Oos} value for the RP with an AAD of 0.25% and particularly for the RP with an AAD of 0.175%.

This completes our analysis of the performance of our different RPs. The final choice of a single "optimal" RP depends on the preferences of the decision-maker. As we have demonstrated, he or she faces a trade-off between the different quality criteria. The final decision will require a weighting of these criteria.

3.5.3 Performance in GAMS

In this final part of our analysis of the RP model, we report on the performance of our GAMS computer implementation (see Appendix 3.B for details) of the two linear formulations of the model. Table 3.6 lists the GAMS model statistics for the two linear models. The model statistics reflect

Table 3.6: Comparison of GAMS Model Statistics.

	Linear model 1	Linear model 2
Blocks of equations	50	42
Single equations	156,026	104,026
Blocks of variables	33	41
Single variables	105,839	157,839
Nonzero elements	3,582,791	2,574,341

This table reports GAMS model statistics with the CPLEX solver for the two linear models using a desktop computer with Windows 7, 64 bit, an Intel Core i5-2400 CPU processor with 3.10 GHz, and 8 GB of RAM.

the advantages and disadvantages of each reformulation. The first linear model has 52,000 fewer variables than the second model (a single auxiliary variable for each scenario in every scenario set for every year) while the second model has 52,000 fewer equations than the first model. This much smaller number of equations leads to more than a million fewer nonzero elements, which is a great strength of the second model.

Table 3.7 reports the running times for both linear models and the three linear programming methods (simplex, interior point, and sifting) we applied to the problems (see Appendix 3.B for details on the GAMS implementation). Not surprisingly, compared to the unconstrained case the introduction of the additional constraints (3.10), (3.11), and (3.12) leads to a significant increase in running times for both linear models and all three linear programming methods. Figure 3.12

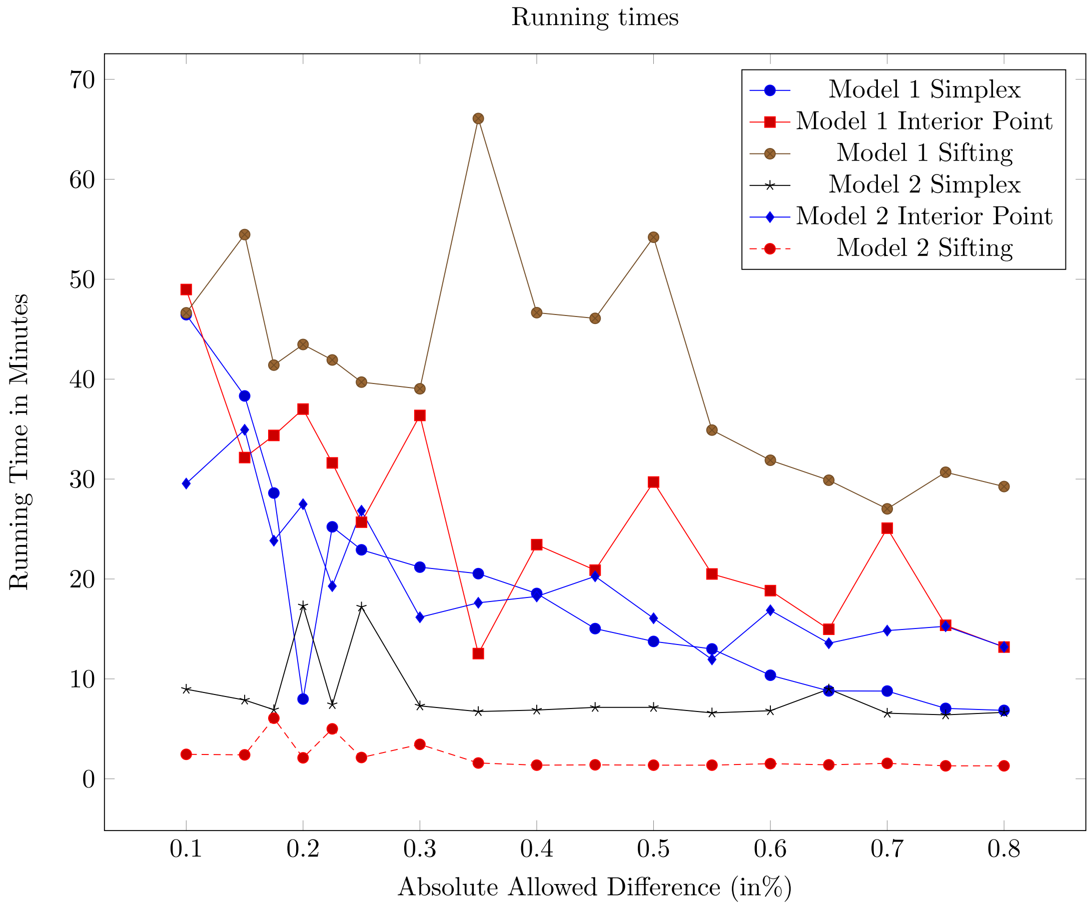
Table 3.7: GAMS Running Times in mm:ss.

	Linear model 1			Linear model 2		
AAD	Simplex	Interior point	Sifting	Simplex	Interior point	Sifting
UC	1:14	0:37	2:43	0:43	0:34	0:36
0.8%	6:51	13:11	29:15	6:39	13:12	1:18
0.75%	7:03	15:22	30:41	6:24	15:16	1:18
0.7%	8:47	25:05	27:01	6:34	14:50	1:33
0.65%	8:48	14:58	29:53	8:59	13:34	1:21
0.6%	10:22	18:50	31:53	6:49	16:52	1:31
0.55%	13:00	20:30	34:54	6:36	11:57	1:22
0.5%	13:45	29:42	54:12	7:09	16:04	1:22
0.45%	15:02	20:53	46:05	7:09	20:16	1:24
0.4%	18:33	23:26	46:39	6:53	18:15	1:22
0.35%	20:32	12:32	66:05	6:44	17:37	1:35
0.3%	21:11	36:22	39:02	7:18	16:10	3:27
0.25%	22:55	25:41	39:42	17:12	26:50	2:08
0.225%	25:13	31:37	41:55	7:26	19:17	5:00
0.2%	7:59	37:00	43:28	17:19	27:29	2:06
0.175%	28:36	34:22	41:24	6:55	23:50	6:04
0.15%	38:19	32:09	54:28	7:53	34:56	2:24
0.1%	46:27	48:58	46:38	8:58	29:33	2:27

This table reports GAMS running times in mm:ss including the model setup with the CPLEX solver in minutes for the two linear models and different lp methods and AADs from unconstrained (UC) to 0.1 using a desktop computer with Windows 7, 64 bit, an Intel Core i5-2400 CPU processor with 3.10 GHz, and 8 GB of RAM.

provides a graphical representation of the results in Table 3.7, which helps us to quickly identify some interesting results. First, the running times differ quite drastically among the two different models and the three different solution methods. In addition, with a few exceptions the running times for both models and each method tend to increase as the constraints on the AADs are tightened. Second,

Figure 3.12: Running Times (in Minutes).



Running times for both models using the simplex, the interior point, and the sifting method, respectively, for different AADs.

the first model has, with a few exceptions, higher running times than the second model for every method. (The exceptions occur for the simplex method at an AAD of 0.2% and for the interior point method at AADs of 0.8%, 0.35%, 0.25%, and 0.15%, respectively.) Third, the running times of the simplex method tend to be shorter than the running times of the interior point method for both models, with a few exceptions for the first model. Fourth, and the most relevant insight, the sifting method delivers for the second model significantly shorter running times for all AAD than any other method for both models. In stark contrast, the sifting method when applied to the first linear model delivers the worst running times across all models and methods. This observation is not surprising since sifting was developed for models with a large ratio of the number of columns (variables) to the number of rows (constraints) in the constraint matrix. Therefore, the sifting method is much better suited to the second model than to the first one.

In sum, the analysis of the running times strongly suggests, for large versions of our RP optimization problem, that solving the second linear reformulation with the sifting method offers the best performance.

3.6 Conclusion

In this paper, we have developed a linear modeling approach for the selection of replicating portfolios in the life insurance industry. The model minimizes the L_1 error between an insurance company's discounted liability cash flows and the discounted RP cash flows. To be useful for practitioners, replicating portfolios must satisfy a variety of different quality criteria. Therefore, we imposed additional constraints on the RP optimization model. In addition, we introduced several tests for an ex post evaluation of the optimal solution to the RP model.

We implemented the model using real-life data sets and demonstrated that our approach can quickly solve realistic versions of the RP model. A detailed sensitivity analysis showed us the trade-offs between the different quality criteria—those imposed as constraints for the in-sample RP optimization and those checked in the subsequent out-of-sample tests. We also determined parameter choices for the constraints, leading to RPs that performed well on all quality criteria. (However, for the final selection of a single RP a decision-maker would have to determine weights for the different quality criteria. That problem of choice is beyond the scope of the present paper.)

Finally, we also performed a detailed performance analysis for the RP optimization model. Solving a variety of model instances with GAMS/CPLEX convincingly demonstrated that the linear reformulation of Feinstein and Thapa (1993) of the L_1 objective, when solved with the sifting method (Bixby et al. (1992)), delivered by far the fastest running times.

To the best of our knowledge, this paper presents the first extensive analysis of an RP optimization model for the life-insurance industry in the operations research literature. We strongly believe that the importance of replicating portfolios in the insurance industry will only continue to grow in the years ahead, when insurance companies will face increased regulatory scrutiny. The analysis in this paper shows that questions relating to replicating portfolios can certainly be a source for stimulating and highly relevant applied work in operations research.

3.A Optimization Models

We present two different linear programming optimization models in this paper.

3.A.1 First Linear Model

For the first model approach we have the following objective function:

$$min_{l,s,z} \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{t=1}^{T} z_{k,c,t} \tag{3.24}$$

We have two sets of constraints to relate the auxiliary variables to the discounted RP cash flows and the discounted liability cash flows:

$$z_{k,c,t} \geq LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \tag{3.25}$$

$$- z_{k,c,t} \leq LCF_{k,c,t} - \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \tag{3.26}$$

We have one set of constraints to limit the notionals, which should hold for a=1,. . . ,A

$$nl(-\sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{1,c,t}/S) \geq \sum_{i=1}^{I_a} (l_{a,i} + s_{a,i}) nv_a \tag{3.27}$$

We have two sets of constraints to limit the scenario set movements, which should hold for set=2,. . . ,Set:

$$\begin{aligned} adub \geq (&\sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i} - s_{a,i}) \\ &- \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{1,c,t}) / \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.28}$$

$$\begin{aligned} adlb \leq (&\sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{a=1}^{A} \sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i} - s_{a,i}) \\ &- \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{1,c,t}) / \sum_{c=1}^{C} \sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.29}$$

Furthermore, the long positions $l_{a,i}$ and the short positions $s_{a,i}$ have to be nonnegative

$$l_{a,i}, s_{a,i} \geq 0, \forall a \forall i \tag{3.30}$$

3.A.2 Second Linear Model

For the second model approach we have the following objective function:

$$min_{l,s,y^+,y^-} \sum_{k=1}^{K} \sum_{c=1}^{C} \sum_{t=1}^{T} (y^+_{k,c,t} + y^-_{k,c,t}) \tag{3.31}$$

In contrast to the first model we only have one set of constraints to relate the auxiliary variables to the discounted RP cash flows and the discounted liability cash flows:

$$y^{+}_{k,c,t} - y^{-}_{k,c,t} = LCF_{k,c,t} - \sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) \tag{3.32}$$

$$y^{+}_{k,c,t}, y^{-}_{k,c,t} \geq 0 \tag{3.33}$$

We face the same set of constraints to limit the notionals as in the first linear model, which should hold for a=1,...,A

$$nl(\sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t}/S) \geq \sum_{i=1}^{I_a}(l_{a,i} + s_{a,i})nv_a \tag{3.34}$$

The two sets of constraints to limit the scenario set movements, which should hold for set=2,...,Set, are also the same as in the first linear model:

$$\begin{aligned} adub \geq (&\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i} - s_{a,i}) \\ &- \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t}) / \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.35}$$

$$\begin{aligned} adlb \leq (&\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{k,c,t,a,i}(l_{a,i} - s_{a,i}) - \sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{a=1}^{A}\sum_{i=1}^{I_a} ACF_{1,c,t,a,i}(l_{a,i} - s_{a,i}) \\ &- \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{k,c,t} + \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t}) / \sum_{c=1}^{C}\sum_{t=1}^{T} LCF_{1,c,t} \end{aligned} \tag{3.36}$$

For the second linear model the long positions $l_{a,i}$ and the short positions $s_{a,i}$ also have to be nonnegative, as was the case in the first linear model

$$l_{a,i}, s_{a,i} \geq 0, \forall a \forall i \tag{3.37}$$

3.B Implementation in GAMS

We implement our optimization models using GAMS. Firstly, we prepare the provided input data using Excel. Secondly, we formulate and solve our optimization problem in GAMS and export our results to Excel. And finally, we use Excel to test the quality of our results.

Preparing the Input Data

The input data sets for the discounted liability cash flows and the discounted cash flows of the candidate assets have been provided by Zurich. We have one Excel sheet including 52,000 (40 years x 1,300 per year) rows and 919 columns for the discounted asset cash flows and one Excel sheet including 52,000 rows and one column for the discounted liability cash flows. We are not able to import Excel sheets of this size into GAMS. Therefore we split the data for the discounted asset cash flows into eight Excel spreadsheets. Each sheet contains five years and the 115 (114 for the first five years) candidate assets used in these years. This results in sheets containing 6,500 rows and 115 or 114 columns. We also split the data for the discounted liability cash flows into eight sheets. Each sheet contains the discounted liability cash flows for 5 years and has one row and 6,500 columns. We import these 16 sheets into GAMS.

Implementation in GAMS

We create two GAMS files to implement the two linear models introduced in Section 3.2. The files are very similar, so we describe them together and mention differences within this description.

Our formulation in GAMS slightly differs from the model instance formulation provided in Section 3.4. We combine the 13 scenario sets with 100 scenarios each for 40 years to 52,000 scenarios. We also combine the six asset classes with up to 720 assets to 919 assets.

As a first step in GAMS we create eight sets for the scenarios and eight sets for the assets (we have to create one set for each Excel sheet). Afterward we create subsets separating the pay fixed swaptions from the other asset classes. We also introduce subsets for all of the 13 scenario sets in the optimization model.

Our parameters for the discounted assets and liability cash flows are imported from Excel and we set parameter values for the AADs. We have decision variables for the long and short asset weights with a nonnegative value and auxiliary variables. The number of decision variables in the second model is thereby twice that of the first model. To improve the clarity of the file we also introduce some auxiliary variables for the replicating portfolio and the notional values of the pay fixed swaptions. Finally we create a variable for our objective value.

We formulate several sets of equations. One equation to calculate the objective value and 25 sets of equations for our constraints. Furthermore we introduce sets of equations to relate the auxiliary variables to the discounted asset and liability cash flows. There are twice as many of these equations in the first model as in the second.

In the next step we solve our model in a loop with just one iteration using the linear programming CPLEX solver. Thereby we create an option file to select a linear programming method. We solve both linear models using the (dual) simplex, the (barrier) interior point, and the sifting method. (For more information about the simplex and the interior point method see Nocedal and Wright (2006). Sifting is a kind of a column generation method that is suitable for models with a large ratio of the number of columns to the number of rows. For more information about sifting, see Bixby et al. (1992).) Then we calculate the trimmed solution of the components of our optimal solutions according to (3.13) and (3.14). In a last step we export the optimal values for our decision variables to Excel. We have to use eight different Excel spreadsheets due to the large number of decision variables.

Testing the Results

We merge the eight Excel files containing the optimal values for the candidate assets into one file. We then subtract the value for the short position of a candidate asset from the value of the long position of that candidate asset. We then insert the optimal values for our candidate assets into two different Excel testing sheets. We use the first sheet to calculate the Oos scenario set movements and the second sheet to calculate the market value movements. We repeat the optimization process in GAMS and the testing in Excel for several AADs.

Part III

Bibliography and Curriculum Vitae

Bibliography

Allaj, E. (2013). The Black-Litterman model: A consistent estimation of the parameter tau. *Financial Markets and Portfolio Management 27*, 217–251.

Anatolyev, S. (2009). Multi-market direction-of-change modeling using dependence ratios. *Studies in Nonlinear Dynamics and Econometrics 13*(1), 232–245.

Anatolyev, S. and A. Gerko (2005). A trading approach to testing for predictability. *Journal of Business and Economic Statistics 23*(4), 455–461.

Barndorff-Nielsen, O. and P. Blaesild (1981). Hyperbolic distributions and ramifications: Contribution to theory and applications. In C. Taillie, G. P. Patil, and B. A. Baldessari (Eds.), *Statistical distributions in scientific work*, pp. 19–44. Dordrecht, Holland: D. Reidel Publishing Company.

Bektas, S. and Y. Sisman (2010). The comparison of L_1 and L_2 minimization methods. *International Journal of the Physical Sciences, September 2010 5(11)*, 1721–1727.

Bertsimas, D., V. Gupta, and I. Paschalidis (2012). Inverse optimization: A new perspective on the Black-Litterman model. *Operations Research 60*(6), 1389–1403.

Best, M. J. and R. R. Grauer (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies 4*, 315–342.

Bevan, A. and K. Winkelmann (1998). Using the Black-Litterman global asset allocation model: Three years of practical experience. *Goldman Sachs Fixed Income Research*.

Bixby, R. E., J. W. Gregory, I. J. Lustig, R. E. Marsten, and D. F. Shanno (1992). Very large-scale linear programming: A case study in combining interior point and simplex methods. *Operations Research 40(5)*, 885–897.

Black, F., M. Jensen, and M. Scholes (1972). The capital asset pricing model: Some empirical tests. In M. Jensen (Ed.), *Studies in the Theory of Capital Markets*, pp. 79–121. New York: Praeger Publishers.

Black, F. and R. Litterman (1990). Asset allocation: Combining investors views with market equilibrium. *Goldman Sachs Fixed Income Research*.

Black, F. and R. Litterman (1992). Global portfolio optimization. *Financial Analysts Journal 48*(5), 28–43.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Policital Economy 81*(3), 637–654.

Bloomfield, P. and W. Steiger (1983). Least absolute deviations: Theory, applications, and algorithms. In P. Huber and M. Rosenblatt (Eds.), *Progress in Probability and Statistics*, Volume 6. Boston: Birkhauser.

Boekel, P., L. van Delft, T. Hoshino, R. Ino, C. Reynolds, and H. Verheugen (2009). Replicating portfolios; An introduction: Analysis and illustrations. *Milliman Research Report November 2009*.

Bollerslev, T. (2008). Glossary to ARCH(GARCH). *CREATES Research Paper 2008 49*.

Brooke, A., D. Kendrick, A. Meeraus, and R. Raman (2014). GAMS - A user's guide; Tutorial by Richard E. Rosenthal. *GAMS Development Corporation, Washington DC, March 2014*.

Burmeister, C. and H. Mausser (2009). Using trading restrictions in replicating portfolios. *Life & Pensions Magazine*, 36–40.

Burmeister, C., H. Mausser, and O. Romanko (2010). Using trading costs to construct better replicating portfolios. *Enterprise Risk Management Symposium Monograph, Society of Actuaries, April 2010*.

CFO Forum (2012). Embedded value. `http://www.cfoforum.nl/embedded_value.html#principles`. Last accessed: 2016-01-05.

Chen, W. and J. Skoglund (2012). Cashflow replication with mismatch constraints. *The Journal of Risk, Summer 2012 14*(4), 115–128.

Cheung, W. (2009). The Black-Litterman model explained. *Journal of Asset Management 11*(4), 229–243.

Chevapatrakul, T. (2013). Return sign forecasts based on conditional risk: Evidence from the UK stock market index. *Journal of Banking & Finance 37*, 2342–2353.

Chopra, V. K. and W. Ziemba (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management 19(2)*, 6–11.

Chow, G., E. Jacquier, M. Kritzman, and K. . Lowry (1999). Optimal portfolios in good times and bad. *Financial Analysts Journal 55*(3), 65–73.

Christoffersen, P. F. and F. X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science 52*(8), 1273–1287.

Christoffersen, P. F., F. X. Diebold, R. S. Mariano, A. S. Tay, and Y. K. Tse (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting 1*(2), 1–22.

Daul, S. and E. Gutiérrez Vidal (2009). Replication of insurance liabilities. *RiskMetrics Journal 9(1)*, 76–96.

De Miguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy. *The Review of Financial Studies 22*(5), 1915–1953.

Devineau, L. and M. Chauvigny (2011). Replicating portfolios: Calibration techniques for the calculation of the Solvency II economic capital. *Bulletin Français d'Actuariat 11*(21), 59–97.

Diebold, F. X. and J. Lopez (1996). Forecast evaluation and combination. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics 14: Statistical Methods in Finance*, pp. 241–268. Amsterdam: Elsevier.

Eberlein, E. and K. Prause (2002). The generalized hyperbolic model: Financial derivatives and risk measures. In H. Geman, D. Madan, S. R. Pliska, and T. Vorst (Eds.), *Mathematical finance – Bachelier Congress 2000*, pp. 245–247. Berlin – Heidelberg: Springer Verlag.

Estrella, A. (1998). A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics 16*, 198–205.

Fabozzi, F. J. (2008). *Handbook of Finance*, Volume I-III. Hoboken, New Jersey: John Wiley & Sons, Inc.

Fabozzi, F. J. and G. W. Buetow (2008a). Interest rate swaps. In F. Fabozzi (Ed.), *Handbook of Finance*, Volume I. Hoboken, New Jersey: John Wiley & Sons, Inc.

Fabozzi, F. J. and G. W. Buetow (2008b). Valuing swaptions. In F. Fabozzi (Ed.), *Handbook of Finance*, Volume III. Hoboken, New Jersey: John Wiley & Sons, Inc.

Feinstein, C. D. and M. N. Thapa (1993). A reformulation of a mean-absolute deviation portfolio optimization model. *Management Science 39*(12), 1552–1553.

Fouque, J. P., G. Papanicolaou, and K. R. Sircar (2000). *Derivatives in financial markets with stochastic volatility*. New York, NY: Cambridge University Press.

Ghalanos, A., E. Rossi, and G. Urga (2015). Independent factor autoregressive conditional density model. *Econometrics Reviews 34*(5), 594–616.

Glosten, L., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*(5), 1779–1801.

Haesen, D., W. G. Hallerbach, T. Markwat, and R. Molenaar (2014). Risk parity versus mean-variance: It's all in the views. *Working Paper Investment Research, Robeco Asset Management, December 2014*. Available at SSRN: http://ssrn.com/paper=2478751.

He, G. and R. Litterman (1999). The intuition behind Black-Litterman model portfolios. *Goldman Sachs Investment Management Research Paper*.

Henriksson, R. D. and R. C. Merton (1981). On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *The Journal of Business 54*(4), 513–533.

Idzorek, T. M. (2005). A step-by-step guide to the Black-Litterman model. Incorporating user-specified confidence levels. *Working Paper Zephyr Associates, April 2005*.

Jacobs, B. I., K. N. Levy, and H. M. Markowitz (2005). Portfolio optimization with factors, scenarios, and realistic short positions. *Operations Research 53*(4), 586–599.

Jacobs, B. I., K. N. Levy, and H. M. Markowitz (2006). Trimability and fast optimization of long-short portfolios. *Financial Analysts Journal 62*(2), 36–46.

Jobson, J. D. and B. M. Korkie (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance 36*, 889–908.

Jondeau, E., S. H. Poon, and M. Rockinger (2001). *Financial modeling under non-Gaussian distributions.* London: Springer.

Jorion, P. (2000). Risk management lessons from long-term capital managment. *European Financial Managment 6*(3), 277–300.

Konno, H. and H. Yamazaki (1991). Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science 37*(5), 519–531.

Koursaris, A. (2011). A primer in replicating portfolios. *barrie + hibbert*.

Kritzman, M. and L. Yuanzhen (2010). Skulls, financial turbulence, and risk management. *Financial Analysts Journal 66*(5), 30–41.

Leung, M. T., H. Daouk, and A.-S. Chen (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Financial Forecasting 16*(2), 173–190.

Litterman, B. (2003). Introduction: Why an equilibrium approach? In B. Litterman and the Quantitative Resources Group (Eds.), *Modern Investment Management- An Equilibrium Approach,*. New Jersey: John Wiley & Sons, Inc.

Litterman, B. and the Quantitative Resources Group (2003). *Modern investment management- An equilibrium approach.* New Jersey: John Wiley & Sons, Inc.

Mahalanobis, P. (1927). Analysis of race-mixture in Bengal. *Journal of the Asiatic Society of Bengal 23*, 301–333.

Mankert, C. (2010). *The Black-Litterman model - Towards its use in practice.* Ph.D. thesis, The Royal Institute of Technology, KTH.

Mankert, C. and M. J. Seiler (2011). Mathematical derivations and practical implications for the use of the Black-Litterman model. *Journal of Real Estate Portfolio Management 17*(2), 139–159.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics 3*(4), 303–328.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters 1*, 21–23.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*, 323–361.

Meucci, A. (2010). The Black-Litterman approach: Original model and extensions. In R. Cont (Ed.), *The Encyclopedia of Quantitative Finance*. Wiley. Extended version available at SSRN: http://ssrn.com/abstract=1117574.

Natolski, J. and R. Werner (2014). Mathematical analysis of different approaches for replicating portfolios. *European Actuarial Journal*, 1–25.

Natolski, J. and R. Werner (2015). Improving optimal terminal value replicating portfolios.

Nocedal, J. and S. J. Wright (2006). *Numerical Optimization*. New York, NY: Springer-Verlag.

Nyberg, H. (2011). Forecasting the direction of the U.S. stock market with dynamic binary probit models. *International Journal of Financial Forecasting 27*(2), 561–578.

Nyberg, H. and H. Pönkä (2015). International sign predictability of stock returns: The role of the United States. *CREATES Research Papers 2015 20*.

Oechslin, J., O. Aubry, M. Aellig, A. Kappeli, D. Bronnimann, A. Tandonnet, and G. Valois (2007). Replicating embedded options. *Life & Pensions Risk*, 47–52.

Pelsser, A. (2003). Pricing and hedging guaranteed annuity options via static option replication. *Insurance: Mathematics and Economics 33*(2), 283–296.

Pesaran, M. H. and A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economics Statistics 10*(4), 461–465.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance 50*(4), 1201–1228.

Pesaran, M. H. and A. Timmermann (2009). Testing dependence among serially corellated multicategory variables. *Journal of the American Statistical Association 104*(485), 325–337.

Rice, J. R. and J. S. White (1964). Norms for smoothing and estimation. *SIAM Review 6*(3), 243–256.

Rudolf, M., H.-J. Wolter, and H. Zimmermann (1997). A linear model for tracking error minimization. *Journal of Banking & Finance 23*, 85–103.

Satchell, S. and A. Scowcroft (2000). A demystification of the Black-Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management 1*(2), 138–150.

Schrager, D. (2008). Replicating portfolios for insurance liabilities. *Aenorm 59*, 57–61.

Seemann, A. (2009). Replizierende Portfolios in der Lebensversicherung. *University of Ulm, Preprint Series 2009-02*.

Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management 21*(1), 49–58.

Varnell, E. M. (2011). Economic scenario generators and Solvency II. *British Actuarial Journal 16*(1), 121–159.

Walters, J. (2013). The factor tau in the Black-Litterman model. Available at SSRN: `http://ssrn.com/paper=1701467`.

Walters, J. (2014). The Black-Litterman model in detail. Available at SSRN: `http://ssrn.com/paper=1314585`.

Woerheide, W. and D. Persson (1993). An index of portfolio diversification. *Financial Services Review 2*(2), 73–85.

Curriculum Vitae of Maximilian Adelmann

Personal Information

Citizenship	Germany
Birthday	14. December, 1983 in Heidelberg, Germany

Education

02/2013 – 02/2017	Ph.D. studies in Business Administration, University of Zurich Advisors: Prof. Karl Schmedders, Prof. János Mayer
09/2009 – 09/2012	Master of Arts in Management and Economics at the University of Zurich
10/2004 – 09/2009	Bachelor of Arts in Economics (B.A. HSG) at the University of St.Gallen
08/2006 – 12/2006	Exchange student, Economics, Copenhagen Business School
08/1994 – 07/2003	Abitur/ A-Levels at the Elisabeth-von-Thadden-Schule, Heidelberg

Employment

10/2012–present	Research Assistant, Chair of Quantitative Business Administration, University of Zurich (Prof. Karl Schmedders)
10/2012–present	Teaching Assistant, Chair of Quantitative Business Administration, University of Zurich (Prof. Karl Schmedders)