Optimal portfolio selection with parameter estimation risks

Statistical modeling and empirical applications

A dissertation in partial fulfillment of the requirements for the degree of
Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

submitted to the
Faculty of Business, Economics, and Management Information Systems
University of Regensburg

by
Felix Kircher

2024

Advisors:
Prof. Dr. Daniel Rösch (University of Regensburg)
Prof. Dr. Ralf Kellner (University of Passau)

Date of disputation:
November 25, 2024

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Portfolio theory has a long tradition in finance. In the early years, prior to the 1950s, portfolio selection was primarily based on security analysis and an intuitive understanding of the concept of diversification (see the discussion in Markowitz (1999)). Exemplary works include Graham and Dodd (1934) and Williams (1938), which mainly focus on the intrinsic value, or the expected returns, of securities, and Leavens (1945), which explains the benefits of diversifying a portfolio of securities. Markowitz (1952, 1959) proposes the first formal and systematic theory of optimal portfolio selection. The mean-variance portfolio selection model, for which Markowitz was awarded the 1990 Alfred Nobel Memorial Prize in Economic Sciences, is a cornerstone of modern finance theory. In particular, the capital asset pricing model (see Sharpe (1964)) as a theory of equilibrium prices in financial markets builds on portfolio theory. The portfolio theory according to Markowitz (1952) specifies how an investor with mean-variance preferences should optimally select his or her portfolio. However, in practice it proves difficult to apply the theory effectively. One of the main reasons for this is parameter estimation risk that arises when determining the model's parameters. Developing methods to enhance the applicability of portfolio theory under parameter estimation risk is an active area of research and is important for effective asset management where these methods can be implemented in practice.

The practical relevance is underscored by the strong growth of the asset management industry in recent decades: Total global assets under management have more than doubled since 2010, reaching approximately $100 trillion by the end of 2022, and further growth is expected in the upcoming years.[1] In general, asset managers offer advice, create investment strategies, and maintain or manage investments for their clients with the objective of increasing the assets' value. There are two main types of clients: retail investors and institutional investors, i.e., banks, insurance companies, investment or pension funds. As a result, asset managers have specialized expertise and access to a large pool of funds compared to the individual client, providing them with greater investment and diversification opportunities. Beyond the selection of the set of

[1] In 2022, assets under management fell as a result of negative stock and bond market performance, but net flows remained positive. See BCG (2023) or similar reports on the asset management industry.

desirable assets, effective asset management requires knowledge of the assets' properties as well as an understanding of efficient portfolio structures. Thus, sophisticated and easily accessible methods can help to enhance the efficiency of selected portfolios in practice.

The research presented in this dissertation contributes to improving practical applications of portfolio theory that use empirical data to estimate the model parameters. After a short review of portfolio theory, the following paragraphs provide an introduction to the literature on parameter estimation risks in optimal portfolio selection and the existing methods for dealing with such estimation risks. After that, the three original research papers that comprise this dissertation are introduced.

Portfolio selection and estimation risks: Literature review

The goal of portfolio selection can be described as maximizing the value of an investment portfolio given the constraints of the investment strategy and a tolerated level of risk. Generally, the notion of risk for an investment portfolio depends on the objective and the preferences of the particular investor. Measuring portfolio risk and quantifying the trade-off between risk and return is an essential part of portfolio selection. To date, the most commonly used framework for quantitative portfolio analysis is mean-variance portfolio theory. For a given risk tolerance, mean-variance analysis provides the optimal allocation of capital among assets with different expected returns and risk, where overall risk is measured by the variance of portfolio returns. The mean-variance framework thus explicitly formalizes the risk-return trade-off. Its crucial insight is that portfolio risk not only depends on the variances of the individual asset returns, but also on their covariances. The advantage of the mean-variance framework—compared to general expected utility-based approaches or alternative approaches that consider downside risk measures—is its simplicity, as it yields analytical solutions in its simplest form or otherwise can be solved efficiently via quadratic programming. As argued in Markowitz (2014), the mean-variance approach can be viewed as a reasonably accurate approximation to many expected utility-based approaches.

While, in general, a model for the multivariate return distribution needs to be specified in order to solve the portfolio optimization problem, in mean-variance analysis the asset return distribution is sufficiently specified by the vector of expected returns and the return covariance matrix. In theory, the model parameters are assumed to be known, in reality, however, they are unknown. Therefore, in order to use the portfolio optimization framework in practice, the parameters must be plausibly determined. Without further model assumptions, the standard approach is to estimate the vector of expected returns by the vector of sample means and the covariance matrix by the sample covariance matrix of historical asset returns. However, the distribution of a finite sample of historical asset returns does not perfectly represent the true distribution of future returns. As a result, the estimated parameters are almost surely not equal to the true model parameters. Using the estimated parameters in portfolio optimization introduces estimation risks into the portfolio selection problem. This implies that the estimated portfolio weights, which are used to determine the capital allocation to the individual assets, are no longer optimal with respect to the true distribution of future asset returns.

Early financial literature considers the problem of parameter estimation in the mean-variance framework, see Friend and Vickers (1965), Mao and Särndal (1966), Cohen and Pogue (1967), and Fried (1970), and recognizes estimation risks, see Kalymon (1971), Frankfurter et al. (1971), and Dickinson (1974). For example, Frankfurter et al. (1971) conduct a simulation experiment with three stocks and use sample estimates of the means and covariance matrix for portfolio selection. The performance of the estimated mean-variance optimal portfolios is found to be poor with respect to the true parameters and indistinguishable from portfolios that are classified as inefficient a priori. This finding begins to question the practical usefulness of portfolio optimization when estimation risks are ignored. Similarly, Dickinson (1974) finds high sampling error in the estimates of weights and risk of minimum variance portfolios. Bloomfield et al. (1977) and Jobson and Korkie (1981) report that empirical estimates of optimal portfolio weights obtained from a historical sample of reasonable size are unreliable and that an equally-weighted portfolio is able to outperform the mean-variance optimized portfolios. Adding to that, Jorion (1985) documents a sharp deterioration in portfolio performance outside the estimation period and instability of the estimated optimal portfolio weights for international portfolio selection. This is primarily attributed to unreliable sample estimates of expected returns, while the covariances can be estimated with relative precision. Michaud (1989) refers to mean-variance portfolio optimization as "error maximization" because it overweights stocks with large expected return estimates which are most likely to have large estimation errors. Broadie (1993) and Kan and Smith (2008) find substantial impact of estimation errors that increases with the number of assets, and show that the estimates of the means and standard deviations of optimized portfolios, as well as the entire in-sample efficient frontier, are optimistically biased estimates of the actual expected returns and standard deviations. In a comprehensive empirical out-of-sample analysis, DeMiguel et al. (2009) find that neither the sample mean-variance optimized portfolios nor portfolio rules developed to explicitly account for estimation risks are able to consistently outperform an equally-weighted portfolio. They show analytically that estimation windows of more than 250 years of monthly data are required to achieve a performance advantage via mean-variance portfolio optimization relative to the equally-weighted benchmark portfolio.

In summary, estimation risks pose a significant challenge to quantitative portfolio selection. As the weights of mean-variance optimized portfolios are sensitive to the input parameters, estimation risks in means and covariances lead to unstable, unreliable, and extreme estimated portfolio weights compared to the truly optimal weights. This can be interpreted as a problem of overfitting the estimated portfolio weights to the available data sample, because the in-sample and out-of-sample portfolio performance measures tend to differ significantly. Within the estimation sample, the benefits of diversification are often grossly overstated. This means that the out-of-sample achievable portfolio mean return is overestimated via the in-sample mean, while the out-of-sample portfolio standard deviation is underestimated, i.e., it is perceived to be lower in-sample than it actually is. As a consequence, the in-sample optimized portfolios often deliver poor out-of-sample performance relative to simple non-optimization-based portfolio strategies. In addition, estimated optimal portfolios exhibit substantial leverage and high turnover of the portfolio positions over time. These facets of estimation risks point to the need for further modeling efforts.

The literature has explored various approaches to address the problem of estimation risks in portfolio optimization. One approach is to improve the out-of-sample properties of the model parameter estimators. Early literature employs the empirical Bayes method to explicitly incorporate estimation risk, leading to shrinkage estimators such as the James–Stein estimator for expected returns. Shrinkage estimators typically shrink a sample estimator towards a target estimator in order to exploit the bias-variance trade-off. The shrinkage target is a highly structured estimator with a small number of free parameters, sometimes even a fixed quantity such as the zero vector, which provides a biased estimator with less variance than the sample estimator. Therefore, the shrinkage method reduces variance by introducing some bias. If the shrinkage intensity is chosen suitably, the resulting estimator is more efficient and less sensitive to estimation error than the original sample estimator. With regard to expected returns, Jobson et al. (1979) and Jorion (1985, 1986) propose to use the James–Stein estimator with a grand mean, i.e., the mean of sample mean returns or the mean return of the sample minimum variance portfolio, as the shrinkage target, while Jobson and Korkie (1981) rely on the grand mean as the estimator of expected returns. Frost and Savarino (1986) use the empirical Bayes method with the prior that all securities have identical expected returns, variances, and correlations, which similarly leads to shrinkage toward their grand means. DeMiguel et al. (2013) conduct an extensive study of shrinkage estimators for the first two moments of asset returns and for the optimal portfolio weights, and investigate various calibration criteria for the optimal shrinkage intensities. Black and Litterman (1992) derive expected returns from market weights based on an equilibrium asset pricing model and allow them to be modified based on subjective investor views. Adding to this, Zhou (2009) combines this approach with information from the data, while Tu and Zhou (2010) extend the idea to form objective-based priors in a Bayesian model.

The formulation of an appropriate structural model offers the potential to reduce estimation risks. Firm-specific indicators and characteristics can be used in a cross-sectional model to more accurately predict expected stock returns. This is shown in Haugen and Baker (1996), who perform mean-variance analysis using expected return predictions from a linear model, and in Lewellen (2015) and Gu et al. (2020), who study the predictability of expected returns using a linear model and advanced machine learning methods, respectively. For the covariance matrix, highly structured models such as factor models or models that assume a significant portion of the parameters to be identical, as explored in Chan et al. (1999), can reduce estimation risks. This is especially true for large investment universes where the sample covariance matrix has a large number of parameters. Ledoit and Wolf (2003, 2004a,b) propose linear shrinkage estimators for the covariance matrix using highly structured models as shrinkage targets. Even better performance can be achieved by using the nonlinear shrinkage estimator proposed in Ledoit and Wolf (2017), as shown in various empirical applications of mean-variance portfolio analysis.

Another approach for dealing with estimation risks works directly on the portfolio weights. The literature suggests that constraints on portfolio weights help to limit estimation errors and can improve the out-of-sample performance of estimated optimal portfolios. This holds even though the constraints impede the ability to construct theoretically optimal portfolios. Frost and Savarino (1988) and Best and Grauer (1991) impose upper bounds and nonnegativity constraints, i.e., no-short-sale constraints, on portfolio weights, while Jagannathan and Ma (2003) show that no-short-sale constraints are equivalent to using a specific covariance matrix shrinkage estimator in portfolio optimization. DeMiguel et al. (2009) extend the framework to flexible norm restrictions on the portfolio weight vector and establish a connection to several shrinkage estimators of the covariance matrix. While these papers consider minimum variance portfolios, Ao et al. (2019) use the LASSO, i.e., the least absolute shrinkage and selection operator, to encourage sparsity in constructing mean-variance optimal portfolios. Brandt et al. (2009) introduce the approach of parametric portfolio weights, as used in DeMiguel et al. (2020), which considers firm-specific characteristics to directly determine the individual portfolio weights. In this approach, the optimization problem is formulated with respect to the parameters of the function specified for the portfolio weights.

Kan and Zhou (2007), Tu and Zhou (2011), and Kan et al. (2022) propose shrinkage approaches for the portfolio weights. Here, the determination of optimal portfolio weights is understood as a statistical estimation problem, where the objective is to minimize the expected utility loss of using an estimator instead of the theoretically optimal portfolio weights. Selected shrinkage targets are the minimum variance or equally-weighted portfolio weights. This approach exploits the bias-variance trade-off because the estimated mean-variance optimal portfolio weights based on the sample estimators of expected returns and covariance matrix have high variance and low bias. By contrast, the target portfolio weights, viewed as an estimator of the theoretically optimal portfolio weights, have lower variance, i.e., even zero variance in the case of the equally-weighted portfolio, but larger bias. With an adequate calibration of the shrinkage intensity, the reduction in variance outweighs the increase in bias, implying that the resulting shrinkage estimator is a more efficient estimator of the optimal portfolio weights.

Finally, another modification of the classical portfolio optimization problem results from the use of robust estimation methods. Garlappi et al. (2007) incorporate an investor's aversion to uncertainty about expected returns by maximizing the portfolio's worst-case performance when the expected returns are within a specified confidence interval around their estimated values. DeMiguel and Nogales (2009) use methods from robust statistics, in particular M- and S-estimators, to formulate the optimization problem in terms of robust risk estimators to achieve greater stability of the estimated optimal portfolio weights.

Contributions of this dissertation

This dissertation contains three original research papers that address key challenges in the practical implementation of mean-variance portfolio theory. Specifically, this dissertation contributes to the understanding of parameter estimation risks in portfolio analysis and develops practical methods to improve the empirical performance of estimated optimal portfolios based on mean-variance analysis. Each of the three research papers has a distinct focus. The first paper uses asset return data in a novel shrinkage approach for estimating the optimal portfolio weights for different asset classes and stock segments. The second paper constructs portfolios of individual stocks by using firm-specific characteristics to estimate expected returns via machine learning methods and by applying a shrinkage approach to the covariance matrix. The third paper proposes a new approach for estimating large-dimensional time-varying covariance matrices that can be used directly for portfolio optimization. Taken together, the research papers expand the set of tools available to academics and practitioners for the application of portfolio optimization models. The following paragraphs provide a summary of the three papers with their distinct approaches and contributions to the literature.

Research paper 1: A shrinkage approach for Sharpe ratio optimal portfolios with estimation risks

The Sharpe ratio—the ratio of the expected excess return to the standard deviation of an investment's return—is among the most important performance measures in academia and practice. It is often used in the ex-post evaluation of portfolio strategies. For this reason, this paper assumes that the investor's objective is to maximize the Sharpe ratio of the realized out-of-sample portfolio returns. Since in practice optimal portfolio weights have to be estimated based on a finite sample of historical returns, they are subject to estimation error. To minimize estimation error, the paper considers a shrinkage estimator for the optimal portfolio weights. The proposed approach takes the entire in-sample estimated efficient frontier as the set of potential shrinkage portfolios. Intuitively, when estimation risks are neglected, the optimal portfolio corresponds to the standard in-sample tangency portfolio. However, when estimation risks are taken into account, it is beneficial to shift more towards the minimum variance portfolio, which only requires the covariance matrix to be estimated. The resulting optimal portfolio is thus the portfolio on the in-sample efficient frontier that adequately balances the exposure to estimation risk in expected returns.

To determine the optimal shrinkage intensity, the paper proposes a novel calibration criterion. The objective is to maximize the expected out-of-sample Sharpe ratio to take into account estimation risk due to sampling variation in the estimated portfolio weights. The paper develops a bootstrap approach to solve the optimization problem and to obtain the shrinkage estimator of the optimal portfolio weights. The proposed bootstrap-based estimator has the advantage of providing flexibility with respect to non-normality of the multivariate return distribution, the inclusion of shrinkage estimators of the covariance matrix, or the modeling of time series dependencies in the data. When implementing these portfolio strategies in practice, it is important to consider the amount of portfolio turnover produced, because transaction costs can be significant and diminish realized Sharpe ratios. In this respect, the paper further contributes by incorporating transaction costs ex-ante into the estimation of optimal portfolio weights in a setting that accounts for parameter estimation risk. In the case where a risk-free asset is available, the paper extends the portfolio optimization problem by including a constraint for the expected out-of-sample portfolio volatility. This yields a two-fund separation, with one fund holding the estimated Sharpe ratio optimal portfolio of risky assets and the other fund holding the risk-free asset.

The paper demonstrates the effectiveness of the proposed methods in a simulation study and an empirical analysis using a variety of datasets, which mainly reflect various segments of the global stock market, but also consider different asset classes. The proposed portfolio weight estimator, which uses a shrinkage covariance matrix estimator and thereby yields a specific combination portfolio of the in-sample tangency, the in-sample minimum variance, and the equally-weighted portfolio, is able to add value to and outperform alternative or non-optimization-based portfolio diversification approaches from the literature. The conducted (pseudo) out-of-sample analysis shows improvements in the realized Sharpe ratios net of transaction costs of about 40% compared to the commonly considered minimum variance and equally-weighted portfolios.

Research paper 2: Estimation of large mean-variance efficient portfolios using machine learning methods

How to best estimate mean-variance optimal portfolios under different circumstances is an open research question. This paper adds to the literature on estimating optimal portfolios when the number of individual stocks is several hundred or more. In this context, firm-specific characteristics can be used as an additional source of information to distinguish the properties and behavior of different stocks. While the approach of parametric portfolio weights, as proposed in Brandt et al. (2009), uses firm characteristics to model the portfolio weights directly, this paper instead shows that it is beneficial to separately handle the steps of first modeling the moments of individual stock returns and second forming optimal portfolios.

The approach proposed in the paper is based on using specialized methods to predict expected returns via firm characteristics and applying an appropriate shrinkage estimator of the return covariance matrix, thereby taking advantage of the flexibility in separately modeling the moments of individual stock returns. The specification of a cross-sectional model for expected returns reduces estimation risk, especially when the number of characteristics is substantially smaller than the number of stocks. In the paper, advanced algorithmic models such as machine learning methods, i.e., linear ridge regression, regression tree ensembles, and artificial neural networks, are used to estimate the expected return model. In particular, regression trees and neural networks are able to capture complex nonlinear and interaction effects of the firm characteristics in relation to expected returns. In the paper, the machine learning methods are specifically tuned to maximize the out-of-sample Sharpe ratio of the estimated mean-variance optimal portfolios.

This paper contributes by extending approaches for incorporating firm characteristics and using machine learning methods in a portfolio optimization framework. The method for constructing large mean-variance optimal portfolios outperforms alternative approaches from the literature in an empirical out-of-sample analysis considering the constituents of three broad US and European stock market indices. To build optimal portfolios under a risk constraint and to satisfy this constraint out-of-sample, the paper extends the mean-variance optimization approach by considering estimation risks of the optimized portfolio weights. Over a period of more than 20 years, the proposed methodology yields realized out-of-sample Sharpe ratios of 2.3 to 2.9 (annualized), outperforming parametric portfolio weights by more than 50 percent and market indices by more than 250 percent. This indicates that there is substantial economic value in using firm characteristics for estimating large mean-variance optimal portfolios. The study finds that a combination of machine learning methods yields improvements of about 10% compared to a linear regression model. The out-of-sample analysis includes strategies that limit turnover through partial rebalancing of the portfolio weights. The results suggest that a substantial reduction in portfolio turnover can be achieved, with the turnover-constrained strategies achieving comparable out-of-sample Sharpe ratios and significantly lower out-of-sample volatilities.

Research paper 3: Large dynamic covariance matrices and portfolio selection with a heterogeneous autoregressive model

A well-established stylized fact about financial asset returns is that their second moments, i.e., their variances and covariances, are time-varying and strongly persistent. This suggests that improvements for portfolio selection can be obtained by considering a dynamic model of the covariance matrix instead of a standard static model. When the number of assets in the portfolio is large, say in the hundreds or thousands, the asset return covariance matrix has a large number of free parameters that must be estimated simultaneously. In this setting, where the matrix dimension is not negligible compared to the number of return observations, the usual sample covariance matrix estimator is subject to large estimation errors. Thus, for large portfolio optimization, the dynamic covariance matrix model must not only adequately represent the covariance dynamics, but also be able to deal with parameter estimation risks.

The paper proposes a new framework for modeling large dynamic covariance matrices based on multiple heterogeneous autoregressive volatility and correlation components. The approach builds on the heterogeneous autoregressive model of Corsi (2009) and its refined version from Bollerslev et al. (2018), which are widely used for realized volatility modeling. In the model, the volatility and correlation components are pre-specified to represent short- and long-term investor horizons and linearly combined to produce predictions of future monthly volatilities and correlations based on daily data. These features make the proposed model flexible and parsimonious, as well as simple to implement and estimate via least squares methods. The paper contributes to the literature by extending the heterogeneous autoregressive approach to the multivariate modeling of dynamic correlations. In addition, the inclusion of state-of-the-art shrinkage estimators for the covariance matrix renders the proposed model suitable for high-dimensional applications with more than a thousand assets.

The paper shows that the proposed modeling framework compares favorably with common benchmark models, such as the static covariance matrix model or multivariate (Realized) GARCH models. For a meaningful comparison of different covariance matrix models, the paper performs a comprehensive empirical out-of-sample analysis based on minimum variance portfolios, which require only the covariance matrix as input, and thus their out-of-sample risk can be readily compared. For different types of constructed portfolios, e.g., based on predefined portfolios or individual stocks, and sizes of ten to 1500 assets, the new model yields lower out-of-sample portfolio volatilities compared to the benchmark models. These differences in portfolio volatilities are found to be statistically significant and economically meaningful. For example, the annualized portfolio standard deviation improves to 8.92% using the proposed model relative to 9.75%, 10.43%, or 11.06% with a Realized DCC model based on Bollerslev et al. (2020), the DCC-NL model from Engle et al. (2019), or a static covariance matrix model, respectively. The paper shows that the proposed methodology adds value to the estimation of large dynamic covariance matrices, achieving a reduction in risk for estimated minimum variance and mean-variance optimal portfolios as examined in the empirical application.

Structure of the dissertation

The remainder of this dissertation is organized as follows.[2] Chapter 2 contains the first paper presenting a novel shrinkage approach for estimating mean-variance optimal portfolio weights with the goal of maximizing the out-of-sample Sharpe ratio. Chapter 3 contains the second paper examining the economic value of using firm characteristics and advanced machine learning methods for portfolio optimization. Chapter 4 contains the third paper presenting a new and easy-to-implement framework for modeling large dynamic covariance matrices that can be readily used in the context of portfolio optimization. Chapter 5 concludes the dissertation by summarizing its main findings and contributions as well as outlining possibilities for extensions and further research.

[2] At the beginning of each chapter, information about the respective paper and its current status is provided. Because the papers were submitted to journals with different requirements and at different times, there may be minor formal differences between the chapters in this thesis.

Chapter 2

A shrinkage approach for Sharpe ratio optimal portfolios with estimation risks

The content of this chapter is joint work with Daniel Rösch[1] and is published as:
Kircher, F. and Rösch, D. (2021). A shrinkage approach for Sharpe ratio optimal portfolios with estimation risks. *Journal of Banking & Finance 133*, 106281.

Available online at: https://doi.org/10.1016/j.jbankfin.2021.106281.

Abstract
We consider the problem of maximizing the out-of-sample Sharpe ratio when portfolio weights have to be estimated. We apply an improved bootstrap-based estimator, and an approximative estimator derived from a Taylor series. In a simulation study and empirical analysis with 15 datasets the proposed estimators outperform the minimum variance and equally weighted portfolio strategies. Out-of-sample Sharpe ratios improve by 15 and 32 percent on average, respectively, in the empirical analysis. While effectively dealing with estimation risks, the estimators produce considerable amounts of turnover. Realized net Sharpe ratios improve by 40 percent on average when the effects of accruing transaction costs are incorporated ex-ante into estimation of portfolio weights. When adding a risk-free asset, net Sharpe ratios remain of the same magnitude and portfolio volatility does not exceed a predefined target level.

Keywords: portfolio optimization; estimation risks; Sharpe ratio; transaction costs; risk constraint.

JEL classification: G11, C13, C63.

[1] Chair of Statistics and Risk Management, Faculty of Business, Economics and Management Information Systems, University of Regensburg, 93040 Regensburg, Germany. Email: daniel.roesch@ur.de.

2.1 Introduction

Mean-variance portfolio theory following Markowitz (1952) requires the specification of assets' expected returns and covariance matrix. In practice, expected returns and covariance matrix are unknown and have to be estimated based on historical data. As a result, parameter estimation risks are a major challenge for practical applications of mean-variance portfolio optimization. Early studies recognize the problem with parameter estimation providing frameworks for dealing with estimation (e.g., Friend and Vickers (1965), Mao and Särndal (1966), Cohen and Pogue (1967) and Fried (1970)) and estimation risks (Kalymon (1971)) in portfolio optimization. A vast number of following studies, e.g., Frankfurter et al. (1971), Jobson and Korkie (1981), Jorion (1985, 1986), Frost and Savarino (1986, 1988), Michaud (1989), and Broadie (1993), find that optimized portfolios based on standard sample estimators of expected returns and covariances are unreliable and perform poorly out of the sample. DeMiguel et al. (2009) add that optimized portfolios underperform naive diversification strategies, i.e., an equally weighted portfolio, for typical sample sizes. The poor out-of-sample performance casts doubt on the practical usefulness of mean-variance-based portfolio optimization.

Shrinkage methods have proven effective in dealing with parameter estimation risks. Kan and Zhou (2007), Tu and Zhou (2011), DeMiguel et al. (2013), and Kan et al. (2022) develop portfolio strategies that take parameter estimation risks into account via shrinkage approaches and thereby improve the out-of-sample performance of mean-variance-optimized portfolios.[2] The former papers simultaneously allow for investments into the risk-free asset, the latter paper focuses on the case without a risk-free asset. Analytical expressions for optimal shrinkage parameters are derived by assuming independent normally distributed asset returns and maximizing the expected out-of-sample utility. Therefore an assumption for the coefficient of relative risk aversion in the mean-variance utility maximization problem is required. DeMiguel et al. (2013) consider several other calibration criteria including a Sharpe ratio analogue and employ the bootstrap to estimate optimal shrinkage parameters. Other shrinkage-type approaches add norm constraints to the portfolio optimization problem such as DeMiguel et al. (2009) who focus on the minimum variance portfolio. Olivares-Nadal and DeMiguel (2018) add transaction costs to the mean-variance portfolio optimization problem and calibrate the transaction cost term empirically to deal with estimation risks. Likewise employing a data-driven approach, Basak et al. (2009) use the jackknife to address the problem of in-sample optimism for the

[2] As for other approaches, Kritzman et al. (2010) argue for using longer-term samples to estimate expected returns, Kirby and Ostdiek (2012) consider different volatility and reward-to-risk timing strategies, Behr et al. (2013) employ constraints on portfolio weights within the optimization process, Garlappi et al. (2007) and Ban et al. (2018) employ methods of robust portfolio optimization, Ao et al. (2019) use the LASSO for the optimization of large portfolios.

out-of-sample variance of the estimated minimum variance portfolio.

In the current state of literature we identify three gaps which we address in this paper. First, empirical evaluation of portfolio strategies is often based on the Sharpe ratio. The Sharpe ratio is one of the most commonly employed performance measures in academia, e.g., in the aforementioned papers, and in practice. Assuming an investor's objective is to maximize the out-of-sample Sharpe ratio, the optimal risk aversion coefficient for expected-utility-based shrinkage approaches is unknown ex-ante. We contribute to the literature by considering the risk aversion coefficient as a shrinkage parameter and determining its optimal value by maximizing the expected out-of-sample Sharpe ratio. The expected out-of-sample Sharpe ratio is a new and practically relevant calibration criterion for shrinkage portfolio weight estimators. Our approach is crucially different from the expected utility and Sharpe-ratio-like criteria considered in previous literature. To solve the problem of directly maximizing the expected out-of-sample Sharpe ratio, we propose an improved bootstrap-based and an approximative estimator based on a Taylor series. Second, we link two strands of the literature which (a) develop improved portfolio strategies under estimation risks, evaluate these portfolio strategies empirically, and consider portfolio turnover or transaction costs ex-post, such as DeMiguel et al. (2013) and Kan et al. (2022), (b) implement ex-ante portfolio optimization with transaction costs such as Olivares-Nadal and DeMiguel (2018). In contrast to Olivares-Nadal and DeMiguel (2018), we incorporate transaction costs ex-ante into estimating portfolio weights that already factor in parameter estimation risks. Third, we obtain a two-fund separation for portfolio weights under parameter uncertainty by extending the optimization problem and including a constraint for the expected portfolio volatility. The fund of risky assets is the shrinkage portfolio maximizing the expected out-of-sample Sharpe ratio. The weight in the risk-free asset is obtained by scaling portfolio weights of risky assets such that the preset risk constraint is met. In contrast to Basak et al. (2009), we consider arbitrary frontier portfolios, target out-of-sample volatilities, and focus on out-of-sample Sharpe ratios as our primary performance measures.

Because in practice portfolio weights have to be estimated based on finite samples, it is essential to take their sampling variation into account. This paper considers the entire estimated (in-sample) efficient frontier as a pool of potential shrinkage portfolios. We calibrate the shrinkage parameter to maximize the expected out-of-sample Sharpe ratio of estimated portfolio weights. This is a natural objective because the Sharpe ratio is the primary performance measure in many academic studies. In absence of estimation errors the resulting optimal portfolio is the standard sample-based tangency portfolio. In presence of estimation errors it proves beneficial to shift towards the minimum variance portfolio. The resulting optimal portfolio is the frontier portfolio that controls the exposure to estimation errors in expected returns adequately. We interpret the optimal shrinkage parameter as an implicit effective risk aversion coefficient comprising two components. One is the value that maximizes the Sharpe ratio for the true efficient frontier without parameter uncertainty, the other component is an add-on because of parameter estimation risks. When using an expected utility calibration criterion for shrinkage parameters (e.g., Kan et al. (2022)), a coefficient of relative risk aversion γ has to be specified explicitly. If, however, the investor's objective is to attain the maximum Sharpe ratio, the optimal γ is ex-ante unknown. We find that in practical applications $\gamma = 3$ often constitutes a reasonable choice, however the optimal value of γ depends on the specific dataset (see Section 2.3 for an illustration). Because in this paper γ is treated as a shrinkage parameter which is estimated for each dataset, the proposed approach effectively solves the ex-ante specification issue for γ.

We propose an application of the bootstrap to directly maximize the expected out-of-sample Sharpe ratio for shrinkage portfolios. When applying the standard bootstrap, a problem is that the expected return estimate of the minimum variance portfolio may be negative in finite samples. Together with overly optimistic estimates of the gain in Sharpe ratio by shifting towards the tangency portfolio this results in an inefficient and too optimistic estimator for the shrinkage parameter. We therefore propose an improved bootstrap procedure and suggest the use of a shrinkage estimator for the expected return vector in the bootstrap application. As an alternative approach we derive an approximation of the out-of-sample Sharpe ratio based on a Taylor series. The result is a less complex bootstrap approach, potentially eliminating the need of a bootstrap application when normally distributed asset returns are assumed and standard sample estimators are used. The bootstrap approaches allow for flexibility regarding non-normality of return distributions, the use of shrinkage estimators of the covariance matrix as in Ledoit and Wolf (2004b) or, if considered, time series dependencies in the data.

We demonstrate the effectiveness of the proposed methods in a comprehensive simulation study and empirical analysis based on 15 datasets. The empirical results show that our portfolio weight estimators outperform the sample-based tangency portfolio on all and the minimum variance and equally weighted portfolios on 12 of the 15 datasets. Out-of-sample Sharpe ratios improve by 15 and 32 percent on average compared with the minimum variance and equally weighted portfolios. The results are robust to the introduction of non-negligible transaction costs. While effectively dealing with estimation risks, the portfolio weight estimators produce considerable amounts of turnover diminishing realized out-of-sample net Sharpe ratios. By incorporating accruing transaction costs ex-ante into the estimation procedure we show that our portfolio strategies taking parameter estimation risks into account benefit from further measures for reducing transaction costs. Out-of-sample net Sharpe ratios improve by 0.19 per annum or 40 percent on average. Our empirical results extend to the case including a risk-free asset and portfolio risk limit. The strategies outperform the minimum variance portfolio on 11 and the equally weighted portfolio on 12 datasets with an average improvement in out-of-sample net Sharpe ratios of 41 and 39 percent. Outperformance is statistically significant on the 5% level on 8 (with respect to the minimum variance portfolio) and on 6 datasets (with respect to the equally weighted portfolio), respectively. Potential underperformance of our proposed strategies is statistically insignificant on all respective datasets. The resulting out-of-sample portfolio volatilities are consistent with a prespecified volatility limit of 5% per annum. We thus conclude that portfolio optimization strategies that take parameter estimation risks into account are able to add value to simplified or non-optimization-based diversification approaches.

The remainder of the paper is organized as follows. Section 2.2 introduces the theory and methodology. Section 2.3 conducts a simulation study demonstrating the theoretical performance of the proposed methods. Section 2.4 presents the results of an empirical application to 15 datasets demonstrating the ability of the methods to achieve considerable gains in out-of-sample Sharpe ratios net of transaction costs compared to, e.g., an equally weighted portfolio strategy. Section 2.5 concludes the paper.

2.2 Theory and methodology

In the following, let N be the number of risky assets. We denote by μ the vector of expected excess returns and by Σ the covariance matrix of excess returns (over the risk-free asset) of the considered assets. $w \in \mathbb{R}^N$ denotes a vector of portfolio weights corresponding to the N risky assets. 1_N is a column vector consisting of N ones.

2.2.1 Maximizing the expected out-of-sample Sharpe ratio

The two-fund separation theorem states that a mean-variance efficient portfolio is the optimal combination of the tangency portfolio with the risk-free asset. The tangency portfolio is the portfolio with the maximum Sharpe ratio, i.e.,

$$w_{\text{TP}} = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu} = \arg\max_{w} \frac{w'\mu}{\sqrt{w'\Sigma w}}, \tag{2.1}$$

where the maximum Sharpe ratio is given by $\theta = \sqrt{\mu'\Sigma^{-1}\mu}$. The problem in practice is that the population values μ and Σ are unknown and have to be estimated. Let $\widehat{\mu}$ and $\widehat{\Sigma}$ denote respective estimators of μ and Σ, e.g., its maximum likelihood estimators. Plugging the estimated parameters into (2.1) we obtain the estimated tangency portfolio $\widehat{w}_{\text{TP}}$. Since $\widehat{\mu}$ and $\widehat{\Sigma}$ are random variables, $\widehat{w}_{\text{TP}}$ is a random variable. The Sharpe ratio in the population or the out-of-sample

Sharpe ratio of $\widehat{w}_{\text{TP}}$ is thus random and it holds that

$$\frac{\widehat{w}_{\text{TP}}'\mu}{\sqrt{\widehat{w}_{\text{TP}}'\Sigma\widehat{w}_{\text{TP}}}} < \theta,$$

i.e., an estimated portfolio does not yield the maximum out-of-sample Sharpe ratio because in a finite sample $\widehat{w}_{\text{TP}} \neq w_{\text{TP}}$ with probability one. Depending on the realizations of $\widehat{\mu}$ and $\widehat{\Sigma}$, the out-of-sample Sharpe ratio of the estimated tangency portfolio $\widehat{w}_{\text{TP}}$ may not even be close to θ. Recognizing that portfolio weights can only be estimated with error, a natural objective would be to maximize the expected out-of-sample Sharpe ratio for some estimated portfolio weight vector $\widehat{w}$. The set of considered portfolio weights needs to be parameterized suitably. An idea is to choose $\widehat{w}$ from the estimated (in-sample) efficient frontier which is spanned by the estimated minimum variance portfolio $\widehat{w}_{\text{MINV}}$ and the estimated tangency portfolio $\widehat{w}_{\text{TP}}$. We thus parameterize $\widehat{w}_\gamma$ as

$$\widehat{w}_\gamma = \widehat{w}_{\text{MINV}} + \frac{1_N'\widehat{\Sigma}^{-1}\widehat{\mu}}{\gamma}\left(\widehat{w}_{\text{TP}} - \widehat{w}_{\text{MINV}}\right), \tag{2.2}$$

where

$$\widehat{w}_{\text{MINV}} = \frac{\widehat{\Sigma}^{-1}1_N}{1_N'\widehat{\Sigma}^{-1}1_N} \quad \text{and} \quad \widehat{w}_{\text{TP}} = \frac{\widehat{\Sigma}^{-1}\widehat{\mu}}{1_N'\widehat{\Sigma}^{-1}\widehat{\mu}}.$$

$\gamma > 0$ is a parameter that traces out the in-sample efficient frontier. $\widehat{w}_\gamma$ corresponds to the solution w_γ of the mean-variance portfolio optimization problem dating back to Markowitz (1952), $\max_w \; w'\mu - \frac{\gamma}{2}\, w'\Sigma w$ subject to $w'1_N = 1$, where population moments are replaced by estimated moments and γ is the coefficient of relative risk aversion. The intuition is as follows. The smaller γ, the larger is the exposure to the estimated tangency portfolio. For $\gamma = 1_N'\widehat{\Sigma}^{-1}\widehat{\mu}$ we obtain the estimated tangency portfolio $\widehat{w}_{\text{TP}}$. The estimated tangency portfolio is exposed to estimation errors in the covariance matrix as well as in expected returns. The estimated minimum variance portfolio is only subjected to estimation errors in the covariance matrix. Under parameter uncertainty it proves beneficial to choose a larger γ in order to mitigate estimation errors from mean returns. We thus shrink the estimated tangency portfolio towards the estimated minimum variance portfolio via the parameter γ. But there is a trade-off. Letting $\gamma \longrightarrow \infty$, we reach the estimated minimum variance portfolio, $\widehat{w}_\gamma = \widehat{w}_{\text{MINV}}$, which completely disregards estimated mean returns and may therefore also be suboptimal. Similarly as in Kan and Zhou (2007) and Kan et al. (2022), it should prove optimal to at least use some information from estimated mean returns. We define our objective of maximizing the expected out-of-sample

Sharpe ratio for shrinkage portfolios $\widehat{w}_\gamma$ by the optimization problem

$$\max_\gamma \mathbb{E}\left[\frac{\widehat{w}'_\gamma \mu}{\sqrt{\widehat{w}'_\gamma \Sigma \widehat{w}_\gamma}}\right]. \tag{2.3}$$

The solution γ^* for γ in (2.3) is a function of the population parameters μ and Σ and depends on their properties, but is generally difficult to derive analytically even when assuming independent normally distributed returns. Because the optimal shrinkage parameter γ^* has to be estimated in empirical applications, we propose bootstrap-based estimators of the unknown true γ^* determining the optimal exposure to the estimated tangency portfolio in (2.2).

2.2.2 Estimating the optimal shrinkage parameter γ^*

In principle, the bootstrap enables us to approximate the plug-in estimator for γ^* (using $\widehat{\mu}$ and $\widehat{\Sigma}$ for μ and Σ) arbitrarily well. Employing the bootstrap, we generate resamples from the return data with time series length T used to estimate portfolio weights. We denote the corresponding bootstrap estimators of the portfolio weights $\widehat{w}_\gamma$ by $\widehat{w}^*_{\gamma,b}$ for the b-th bootstrap sample. A bootstrap analogue of (2.3) is then defined by

$$\widehat{\gamma^*} = \arg\max_\gamma \frac{1}{B}\sum_{b=1}^{B} \frac{\widehat{w}^{*,\prime}_{\gamma,b}\, \widehat{\mu}_{\mathrm{a}}}{\sqrt{\widehat{w}^{*,\prime}_{\gamma,b}\, \widehat{\Sigma}\, \widehat{w}^*_{\gamma,b}}}, \tag{2.4}$$

where B is the number of resampling iterations and $\widehat{\mu}_{\mathrm{a}}$ is an estimator of μ. By following the standard approach and using the sample mean estimator $\widehat{\mu}$ in place of $\widehat{\mu}_{\mathrm{a}}$ two problems emerge.

First, the fact that the dispersion of the components of $\widehat{\mu}$ is wider than the dispersion of the population values in μ, i.e., estimated mean returns varying too widely on average, results in an overly optimistic assessment of the Sharpe ratio for investing into the tangency portfolio in comparison to the minimum variance portfolio. We address this problem by using a shrinkage estimator of $\widehat{\mu}$ in the bootstrap procedure. We consider the shrinkage estimator

$$\widehat{\mu}_{\mathrm{sh}} = (1-\alpha)\, \widehat{\mu} + \alpha\, \widetilde{\mu}\, 1_N \tag{2.5}$$

which shrinks $\widehat{\mu}$ towards a scalar random variable $\widetilde{\mu}$.[3] By specifying $\widetilde{\mu} = 1'_N \widehat{\mu}/N$, which is the grand mean of estimated mean returns $\widehat{\mu}$, we reduce the dispersion of the components of $\widehat{\mu}$.

[3] Shrinkage estimators of the mean of this type have first been considered in a portfolio optimization context in Jobson et al. (1979). Other papers are Frost and Savarino (1986) and Jorion (1985, 1986) following Bayesian approaches and DeMiguel et al. (2013). Our calibration criterion differs from the aforementioned papers.

Assuming that $\widehat{\mu} \sim N(\mu, \Sigma/T)$ we can show that

$$\mathbb{E}\left[(\widehat{\mu} - \widetilde{\mu}\,1_N)'\,(\widehat{\mu} - \widetilde{\mu}\,1_N)\right] = \left(\text{trace}(\Sigma) - 1_N'\Sigma\,1_N/N\right) / \,T + (\mu - \bar{\mu}\,1_N)'\,(\mu - \bar{\mu}\,1_N), \quad (2.6)$$

where $\bar{\mu} = 1_N'\mu/N$ denotes the grand mean of μ. (2.6) implies that the dispersion of the components of $\widehat{\mu}$ is on average wider than the dispersion of expected returns μ. Considering that for the expected dispersion of the shrinkage estimator $\widehat{\mu}_{\text{sh}}$ around its mean it holds that

$$\mathbb{E}\left[(\widehat{\mu}_{\text{sh}} - \widetilde{\mu}_{\text{sh}}1_N)'\,(\widehat{\mu}_{\text{sh}} - \widetilde{\mu}_{\text{sh}}1_N)\right] = (1-\alpha)^2\;\mathbb{E}\left[(\widehat{\mu} - \widetilde{\mu}\,1_N)'\,(\widehat{\mu} - \widetilde{\mu}\,1_N)\right]$$

with $\widetilde{\mu}_{\text{sh}} = 1_N'\widehat{\mu}_{\text{sh}}/N = 1_N'\widehat{\mu}/N$, we aim to match the dispersion of the components of μ, i.e., $(\mu - \bar{\mu}\,1_N)'\,(\mu - \bar{\mu}\,1_N)$. This yields the shrinkage intensity

$$\alpha^* = 1 - \sqrt{\frac{(\mu - \bar{\mu}\,1_N)'\,(\mu - \bar{\mu}\,1_N)}{\mathbb{E}\left[(\widehat{\mu} - \widetilde{\mu}\,1_N)'\,(\widehat{\mu} - \widetilde{\mu}\,1_N)\right]}}.$$

Because $\alpha^* > 0$ in general, we obtain a shrinkage effect of $\widehat{\mu}$ towards $\widetilde{\mu}$ via (2.5). To use the shrinkage estimator $\widehat{\mu}_{\text{sh}}$ in practice, we need an estimator $\widehat{\alpha^*}$ for α^*. Our estimator $\widehat{\alpha^*}$ is described in Section 2.A.1 in the appendix.

The second problem arising when using the standard bootstrap is that the scalar random variable $\widehat{c} = 1_N'\widehat{\Sigma}^{-1}\widehat{\mu}$, i.e., the denominator of the estimated tangency portfolio or the ratio of the mean return of the minimum variance portfolio to its variance, turns negative with positive probability. Being especially relevant for small samples, a negative $\widehat{c}$ is theoretically implausible in most practically relevant cases because for its population counterpart it holds that $c = 1_N'\Sigma^{-1}\mu = \mu_{\text{MINV}}/\sigma^2_{\text{MINV}}$. Because the expected return of the population minimum variance portfolio μ_{MINV} is positive (for positive risk exposure), c should also be strictly positive. Otherwise, a negative $\widehat{c}$ or $\widehat{\mu}_{\text{MINV}}$ results in the Sharpe ratio being an improper performance measure for the optimization procedure. If the capital asset pricing model holds, c is (approximately) equal to the relative risk aversion of the representative investor. For the coefficient of relative risk aversion typical values from the literature range from 1 to 10 (see, e.g., Brandt (2010)). Considering the two arguments, it is in most cases neither plausible that c is negative nor that it is too small. To address this problem we explore two alternatives in order to achieve an improved estimator for c. One simple approach introduces a positive lower bound $c_{\min}$ which ensures that the estimator $\widehat{c}_{\min}$ for c cannot take values below the predetermined threshold value. The disadvantage of this simple approach is that for a fixed $c_{\min}$ lower values of $\widehat{c}_{\min}$ are impossible so that $c_{\min}$ has to be chosen carefully. We set $c_{\min} = 3$ in the following. A second and more complex approach to estimate c is to use a penalized maximum likelihood method. This approach translates into a shrinkage estimator $\widehat{c}_{\text{PML}}$ for c taking only positive values but not requiring a strict lower bound on its values. Both estimators, $\widehat{c}_{\min}$ and $\widehat{c}_{\text{PML}}$, are further described in Section 2.A.2 in the appendix.

For the proposed bootstrap procedure in (2.4) we thus employ an adjusted estimator for $\widehat{\mu}$ which addresses the problems described before. That is, we adjust the shrinkage estimator from (2.5) such that

$$\widehat{\mu}_{\text{a}} = \widehat{\mu}_{\text{sh}} + \max\left\{\frac{\widehat{c}_{\text{a}} - 1_N'\widehat{\Sigma}^{-1}\widehat{\mu}_{\text{sh}}}{1_N'\widehat{\Sigma}^{-1}1_N}, 0\right\} 1_N$$

is used in the bootstrap procedure which ensures that $1_N'\widehat{\Sigma}^{-1}\widehat{\mu}_{\text{a}} \geq \widehat{c}_{\text{a}}$, where $\widehat{c}_{\text{a}}$ is either given by $c_{\min}$ or the penalized maximum likelihood estimator $\widehat{c}_{\text{PML}}$. To finally solve (2.4) we specify a grid for possible values of γ and choose the value of γ, $\widehat{\gamma^*}$, that yields the maximum bootstrap mean 'out-of-sample' Sharpe ratio. We then estimate the portfolio weights of the portfolio with maximum expected out-of-sample Sharpe ratio via

$$\widehat{w}_{\widehat{\gamma^*}} = \widehat{w}_{\text{MINV}} + \frac{1_N'\widehat{\Sigma}^{-1}\widehat{\mu}}{\widehat{\gamma^*}}\left(\widehat{w}_{\text{TP}} - \widehat{w}_{\text{MINV}}\right). \tag{2.7}$$

As an alternative and easier to implement bootstrap procedure for determining γ^* we approximate the out-of-sample Sharpe ratio by employing a second-degree Taylor polynomial. That is, we approximate the function $f(w) = \frac{w'\mu}{\sqrt{w'\Sigma w}}$ for w near some prespecified portfolio weight vector w_0 by

$$\widetilde{f}(w, w_0) = \frac{w_0'\mu}{\sqrt{w_0'\Sigma w_0}} + \sum_{i=1}^{N} f_i^1(w_0)\left(w_i - w_{0,i}\right) + \frac{1}{2}\sum_{i,j=1}^{N} f_{i,j}^2(w_0)\left(w_i - w_{0,i}\right)\left(w_j - w_{0,j}\right), \tag{2.8}$$

where

$$f_i^1(w_0) = \frac{\partial f}{\partial w_i}(w_0) = \frac{\mu_i}{\sqrt{w_0'\Sigma w_0}} - \frac{w_0'\mu}{\left(w_0'\Sigma w_0\right)^{3/2}}\left(\Sigma w_0\right)_i, \tag{2.9}$$

$$\begin{aligned} f_{i,j}^2(w_0) &= \frac{\partial^2 f}{\partial w_i \partial w_j}(w_0) \\ &= -\frac{\left(\mu w_0'\Sigma + \Sigma w_0 \mu'\right)_{i,j}}{\left(w_0'\Sigma w_0\right)^{3/2}} + 3\frac{w_0'\mu}{\left(w_0'\Sigma w_0\right)^{5/2}}\left(\Sigma w_0 w_0'\Sigma\right)_{i,j} - \frac{w_0'\mu}{\left(w_0'\Sigma w_0\right)^{3/2}}\left(\Sigma\right)_{i,j}. \end{aligned}$$

Because the aim is to approximate $f(\widehat{w}_\gamma)$, i.e., the out-of-sample Sharpe ratio of estimated portfolio weights $\widehat{w}_\gamma$, we set w_0 to its population counterpart w_γ and obtain as an approximate solution to (2.3)

$$\gamma^* = \arg\max_\gamma \mathbb{E}\left[\frac{\widehat{w}_\gamma'\mu}{\sqrt{\widehat{w}_\gamma'\Sigma\widehat{w}_\gamma}}\right] \approx \arg\max_\gamma \mathbb{E}\left[\widetilde{f}(\widehat{w}_\gamma, w_\gamma)\right]. \tag{2.10}$$

As a result of (2.10) a few terms from (2.9) need to be estimated. These are $w_\gamma'\mu = \mu_{\text{MINV}} + \frac{1}{\gamma}\psi^2$, $w_\gamma'\Sigma w_\gamma = \sigma^2_{\text{MINV}} + \frac{1}{\gamma^2}\psi^2$, and $\Sigma w_\gamma = \sigma^2_{\text{MINV}} 1_N + \frac{1}{\gamma}(\mu - \mu_{\text{MINV}} 1_N)$, where ψ^2 is the difference of the squared Sharpe ratios of the population tangency portfolio and the population minimum variance portfolio given by $\psi^2 = \mu'\Sigma^{-1}\mu - (1_N'\Sigma^{-1}\mu)^2 / 1_N'\Sigma^{-1}1_N$. From (2.8) the expected values $\mathbb{E}\left[\widehat{w}_\gamma - w_\gamma\right]$ and $\mathbb{E}\left[(\widehat{w}_\gamma - w_\gamma)(\widehat{w}_\gamma - w_\gamma)'\right]$ need to be estimated.

We estimate μ_i, $(\Sigma)_{i,j}$ and $\mu_i - \mu_{\text{MINV}}$ from (2.9) by using their maximum likelihood estimators for $\widehat{\mu}$ and $\widehat{\Sigma}$ and estimate σ^2_{MINV} via $\widehat{\sigma}^2_{\text{MINV}} = T/(T-N)\left(1_N'\widehat{\Sigma}^{-1}1_N\right)^{-1}$. The expected return of the minimum variance portfolio $\mu_{\text{MINV}} = c\,\sigma^2_{\text{MINV}}$ is estimated via $\widehat{\mu}_{\text{MINV}} = \widehat{c}_{\text{a}}\,\widehat{\sigma}^2_{\text{MINV}}$, where $\widehat{c}_{\text{a}}$ is either $\widehat{c}_{\text{min}}$ or $\widehat{c}_{\text{PML}}$ from above which ensure a positive estimate. Because the plug-in or sample-based estimator of ψ^2 is heavily upwards biased in small samples, we use the adjusted estimator given in Kan and Zhou (2007). Finally, we estimate $\mathbb{E}\left[\widehat{w}_\gamma - w_\gamma\right]$ and $\mathbb{E}\left[(\widehat{w}_\gamma - w_\gamma)(\widehat{w}_\gamma - w_\gamma)'\right]$ via a standard bootstrap procedure with B replications and solve (2.10) numerically for γ to obtain an estimator $\widehat{\gamma^*}$ which we then use in (2.7).[4]

2.2.3 Adding a risk-free asset

Up to this point the aim was to construct a portfolio containing only risky assets that achieves the maximum expected out-of-sample Sharpe ratio. We now consider the case in which additionally a risk-free asset is available. The objective remains the same as in (2.3), i.e., the expected out-of-sample Sharpe ratio is maximized, but an additional risk constraint is required to hold. This leads to the problem of

$$\max_{\gamma,\lambda} \mathbb{E}\left[\frac{\widehat{w}_{\gamma,\lambda}'\mu}{\sqrt{\widehat{w}_{\gamma,\lambda}'\Sigma\widehat{w}_{\gamma,\lambda}}}\right] \qquad \text{s. t.} \qquad \mathbb{E}\left(\sqrt{\widehat{w}_{\gamma,\lambda}'\Sigma\widehat{w}_{\gamma,\lambda}}\right) \leq \sigma, \tag{2.11}$$

where $\widehat{w}_{\gamma,\lambda} = \lambda\,\widehat{w}_\gamma$ are scaled portfolio weights with $\widehat{w}_\gamma$ as in (2.2). The weights of $\widehat{w}_\gamma$ sum to one so that λ controls the exposure to the fund containing only risky assets and $1-\lambda$ is

[4] To explicitly calculate $\mathbb{E}\left[\widehat{w}_\gamma - w_\gamma\right]$ and $\mathbb{E}\left[(\widehat{w}_\gamma - w_\gamma)(\widehat{w}_\gamma - w_\gamma)'\right]$ the analytical results from Okhrin and Schmid (2006) for the distribution of portfolio weights $\widehat{w}_\gamma$ under the assumption of independent and normally distributed return data can be used, which yield good results, if the bootstrap approach is too time-consuming.

the portfolio weight of the risk-free asset. In (2.11) the expected out-of-sample volatility is constrained to not exceed a predefined target value σ. In this context it is of course also possible to consider the expected out-of-sample variance as in Basak et al. (2009). To solve (2.11) it is useful to note that the Sharpe ratio $\frac{\widehat{w}'_{\gamma,\lambda}\mu}{\sqrt{\widehat{w}'_{\gamma,\lambda}\Sigma\widehat{w}_{\gamma,\lambda}}}$ is independent of λ. Thus, we can first solve (2.11) with respect to γ for the maximum expected out-of-sample Sharpe ratio portfolio as in (2.3) yielding $\widehat{w}_{\gamma^*}$. Secondly, we scale $\widehat{w}_{\gamma^*}$ by $\lambda^* = \sigma / \mathbb{E}\left(\sqrt{\widehat{w}'_{\gamma^*}\Sigma\widehat{w}_{\gamma^*}}\right)$ to arrive at the desired portfolio weights $\widehat{w}_{\gamma^*,\lambda^*} = \lambda^* \, \widehat{w}_{\gamma^*}$ for the N risky assets and $1 - \lambda^*$ for the risk-free asset. To estimate $\widehat{w}_{\gamma^*,\lambda^*}$ we first estimate γ^* by $\widehat{\gamma^*}$ following the procedures from Section 2.2.2. We then estimate $\mathbb{E}\left(\sqrt{\widehat{w}'_{\gamma^*}\Sigma\widehat{w}_{\gamma^*}}\right)$ via repeated 5-fold cross-validation which is a related method to the jackknife used in Basak et al. (2009). From this we calculate $\widehat{\lambda^*}$ and obtain $\widehat{w}_{\widehat{\gamma^*},\widehat{\lambda^*}} = \widehat{\lambda^*} \, \widehat{w}_{\widehat{\gamma^*}}$ as the estimated portfolio weights for the risky assets and $1 - \widehat{\lambda^*}$ as the estimated portfolio weight for the risk-free asset.

2.2.4 Introducing transaction costs

Implementing, for instance, a rolling portfolio strategy in practice means that portfolio weights are updated periodically. The resulting transactions are costly and reduce the portfolio return earned over the investment horizon. Portfolio weights that fluctuate greatly over time due to estimation errors result in high trading costs. Thus, controlling for estimation errors contributes to the reduction of trading costs. Nevertheless, even when implementing the methodology above which directly addresses the problem of parameter uncertainty, some variation induced by the estimation process remains, leading to a higher portfolio turnover than necessary. Theoretically, the portfolio should only be rebalanced if by means of rebalancing a gain in Sharpe ratio after accounting for transaction costs is to be expected. Therefore, transaction costs should optimally already be included ex-ante at portfolio construction, and not only considered ex-post when portfolio performance is evaluated (see, e.g., Füss et al. (2014), Mei et al. (2016)).

Ideally, we would like to include associated trading costs into our optimization of the expected out-of-sample Sharpe ratio. Unfortunately this is not easily feasible for us because it would require resampling all estimated portfolio weights from previous periods, see Füss et al. (2014) for such an application to the minimum variance portfolio. For our case we present a simple alternative approach by considering transaction costs solely on an in-sample basis which, nevertheless, works well empirically as shown in Section 2.4. We follow a two-step approach by firstly estimating γ^* as described in Section 2.2.2. Because γ^* is the solution to (2.3) and $\widehat{w}_{\gamma^*}$ solves $\max_w \; w'\widehat{\mu} - \frac{\gamma^*}{2} w'\widehat{\Sigma}w$ subject to $w'1_N = 1$ (see (2.2)), γ^* specifies the portfolio on the in-sample efficient frontier that maximizes the expected out-of-sample Sharpe ratio and trades off the estimated portfolio mean $w'\widehat{\mu}$ and the estimated portfolio variance $w'\widehat{\Sigma}w$ just right. Proceeding from that we, secondly, use γ^* to find the portfolio that is located on the efficient frontier including associated transactions costs. That is, to consider trading costs we estimate the portfolio weights as in (2.7), but add a penalty term that reflects the amount of trading costs incurred by rebalancing the portfolio in this period:

$$\widehat{w}_{\widehat{\gamma^*}} = \arg\max_{w}\ w'\widehat{\mu} - \frac{\widehat{\gamma^*}}{2}\, w'\widehat{\Sigma}w - \kappa\, |w - w_0|_1 \tag{2.12}$$

subject to $w'1_N = 1$, where $|w - w_0|_1 = \sum_{i=1}^{N} |w_i - w_{0,i}|$ and w_0 is the vector of prevailing portfolio weights. In (2.12) we include constant proportional transaction costs in the amount of κ for all assets. This type of transaction costs can alternatively be interpreted as a turnover constraint and has been considered in a portfolio optimization context in Brodie et al. (2009), Füss et al. (2014), Kourtis (2015) and Olivares-Nadal and DeMiguel (2018). The extent to which rebalancing of the portfolio is economically feasible now depends on the in-sample portfolio mean return after transaction costs given by $w'\widehat{\mu} - \kappa\, |w - w_0|_1$, the in-sample portfolio variance $w'\widehat{\Sigma}w$ and the estimated coefficient $\widehat{\gamma^*}$ which trades off portfolio mean and variance. Setting $\kappa = 0$ we recover the estimated portfolio weights from (2.7) that maximize the expected out-of-sample Sharpe ratio without transaction costs. When additionally a risk-free asset is included and portfolio weights for risky assets are scaled by $\widehat{\lambda^*}$ as described in Section 2.2.3, we first solve the optimization problem

$$\widehat{w}_{\widehat{\gamma^*}} = \arg\max_{w}\ w'\widehat{\mu} - \frac{\widehat{\gamma^*}}{2}\, w'\widehat{\Sigma}w - \kappa\, \left|w - \widehat{\lambda_0^*}/\widehat{\lambda^*}\, w_0\right|_1 \tag{2.13}$$

subject to $w'1_N = 1$, where w_0 are prevailing portfolio weights for the risky assets that sum to one and are then scaled by $\widehat{\lambda_0^*}$. The transaction costs thus amount to $\kappa\, |\widehat{\lambda^*}\, w - \widehat{\lambda_0^*}\, w_0|_1 = \kappa\, \widehat{\lambda^*}\, |w - \widehat{\lambda_0^*}/\widehat{\lambda^*}\, w_0|_1$ which is proportional to $\kappa\, |w - \widehat{\lambda_0^*}/\widehat{\lambda^*}\, w_0|_1$. After solving (2.13) we rescale the resulting portfolio weights by the new scaling factor $\widehat{\lambda^*}$.

2.3 Simulation study

We demonstrate the theoretical performance of the proposed portfolio weight estimators in a simulation study. The simulation study is based on 15 datasets presented in Table 2.1. The listed datasets comprise many of the datasets typically considered in the literature, cf., e.g., DeMiguel et al. (2009), Tu and Zhou (2011), Kirby and Ostdiek (2012), DeMiguel et al. (2013), and Kan et al. (2022). For each dataset we try to use the longest possible time series.

Table 2.1: 15 datasets used in the analyses with number of assets N and considered time period.

#	Dataset	N	Time Period	Source[a]
1	3 Fama-French Factors	3	07/1926 – 12/2019	Ken French
2	4 World Regions (MSCI: North America, Europe, Pacific, EM)	4	01/1988 – 12/2019	Datastream
3	9 Asset Classes (Russell 1000, Russell 2000, Russell 1000 Growth, Russell 1000 Value, USBIG Treasury, USBIG Treasury 20+ year, USBIG Corporate, S&P GSCI Commodity, DJ US Select REIT)	9	01/1980 – 12/2019	Datastream
4	10 Industry Portfolios	10	07/1926 – 12/2019	Ken French
5	10 Momentum Portfolios	10	01/1927 – 12/2019	Ken French
6	10 Portfolios based on Idiosyncratic Volatility	10	07/1963 – 12/2019	Ken French
7	10 Portfolios based on Long-term Reversal	10	01/1931 – 12/2019	Ken French
8	10 Portfolios based on Short-term Reversal	10	07/1926 – 12/2019	Ken French
9	18 Developed Countries (MSCI: USA, Canada, UK, Germany, France, Italy, Spain, Netherlands, Belgium, Switzerland, Austria, Denmark, Sweden, Norway, Japan, Hong Kong, Singapore, Australia)	18	01/1970 – 12/2019	Datastream
10	25 Size/Book-to-Market(BM) Portfolios	25	07/1926 – 12/2019	Ken French
11	25 Operating Profitability(OP)/Investment(INV) Portfolios	25	07/1963 – 12/2019	Ken French
12	30 Industry Portfolios	30	07/1926 – 12/2019	Ken French
13	32 Size/OP/INV Portfolios	32	07/1963 – 12/2019	Ken French
14	Constituents of the S&P 100 as of 02/2020 with a complete return data history	50	02/1973 – 12/2019	Datastream
15	100 Size/BM Portfolios	100	07/1945 – 12/2019	Ken French

[a] Most datasets are from Kenneth R. French's website (also the data on the risk-free rate of return) for which we express gratitude for making them available. The remaining data are from Refinitiv Datastream.

In the simulation study we generate asset (excess) returns independently over time from a multivariate normal distribution. The parameters of the distribution are set to the sample estimates over the full time period of each dataset. We consider the methods from Sections 2.2.1 and 2.2.2 for estimating portfolio weights that maximize the expected out-of-sample Sharpe ratio containing only risky assets without considering transaction costs. Benchmark strategies are the sample tangency portfolio based on the maximum likelihood estimator of the parameters, the sample minimum variance and the equally weighted portfolios. The sample minimum variance and the equally weighted portfolios have been shown to be tough to beat in practice, see DeMiguel et al. (2009). The methods in this paper shrink the sample tangency portfolio, which makes full use of expected return estimates, towards the minimum variance portfolio, which makes no use of expected return estimates. We therefore aim for the right balance in

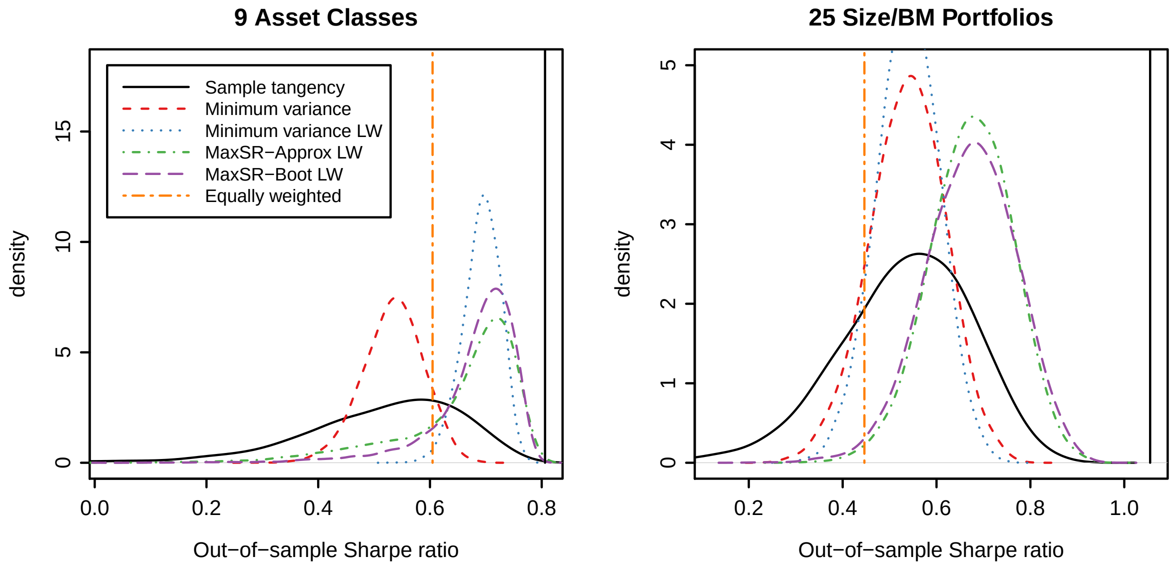

Figure 2.1: Densities of annualized out-of-sample Sharpe ratios based on 10,000 simulation runs for different portfolio strategies (for details cf. Section 2.3). Portfolio strategies are displayed with different colors and line types. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 based on (2.10) and (2.4), respectively, using $\widehat{c}_{\text{PML}}$ from (2.16). The rightmost vertical line represents the maximum attainable Sharpe ratio when the population parameters (expected returns and covariance matrix) are known. Population parameters are based on datasets of 9 Asset Classes (left) and 25 Size/BM portfolios (right), cf. Table 2.1. The length of the estimation period for each portfolio strategy is $T = 120$.

using expected return estimates for portfolio construction under estimation risks. An improved covariance matrix estimator can enhance out-of-sample performance further. The shrinkage estimator of Ledoit and Wolf (2004b), practically speaking, shifts portfolio weights of the minimum variance portfolio towards the equally weighted portfolio. We can thus view our proposed estimators using a shrunk covariance matrix as a combination portfolio of the sample tangency, the sample minimum variance and the equally weighted portfolios. We thereby aim to achieve an improvement over these three typically considered portfolio strategies. We compare our proposed estimators with the estimator from Kan et al. (2022) that is based on maximizing the expected out-of-sample utility and aim to achieve similar or improved expected out-of-sample Sharpe ratios depending on the specific dataset.

We consider lengths of estimation periods T between 60 and 600 months which are common values in the literature (cf. Kan and Zhou (2007), Ao et al. (2019), and Kan et al. (2022)). The number of bootstrap samples that we use to estimate γ^* as described in Section 2.2.2 is set to

$B = 1000$. We evaluate the out-of-sample performance of portfolio strategies via the prespecified vector of expected returns and covariance matrix, i.e., the population parameters based on each dataset. Each single simulation run yields an out-of-sample Sharpe ratio for each portfolio strategy whereby portfolio weights are estimated based on the generated return data with time series length T. We run the simulation with $T = 120$ ten thousand times and present density plots of the 10,000 annualized out-of-sample Sharpe ratios for six portfolio strategies and two parameter constellations (based on the respective datasets) in Figure 2.1. The left-hand plot shows a much larger variance of the distribution of the out-of-sample Sharpe ratio of the sample tangency portfolio (solid) as compared with the minimum variance portfolio (short-dashed). The shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) further reduces the variance of the out-of-sample Sharpe ratio of the estimated minimum variance portfolio and improves its performance. The shrinkage minimum variance portfolio (dotted) displays superior out-of-sample Sharpe ratios compared to the sample tangency portfolio. Our proposed approximative (dot-dashed) and bootstrap-based (long-dashed) estimators shrink the weights of the sample tangency towards the minimum variance portfolio. Its Sharpe ratio distributions approach the Sharpe ratio distribution of the minimum variance portfolio. Most of the probability mass is located above the Sharpe ratio of the equally weighted portfolio. In this parameter constellation, the shrinkage minimum variance portfolio achieves relatively high out-of-sample Sharpe ratios. Its mean out-of-sample Sharpe ratio is 0.69 while the theoretically optimal portfolio based on population parameters exhibits a Sharpe ratio of 0.81. The approximative and bootstrap-based estimators achieve mean out-of-sample Sharpe ratios of 0.65 and 0.68, respectively, exceeding the Sharpe ratio of the equally weighted portfolio of 0.60 and approaching the Sharpe ratio of the minimum variance portfolio. The right-hand plot shows a substantial shrinkage effect of our proposed strategies. On this dataset, the mean out-of-sample Sharpe ratios of the sample tangency and the minimum variance portfolios are 0.49 and 0.54, respectively, which our approximative and bootstrap-based estimators improve upon by obtaining mean out-of-sample Sharpe ratios of 0.67.

Figure 2.2 depicts the mean out-of-sample Sharpe ratios of portfolio weight estimators for increasing monthly time series lengths T from 60 to 600. Based on 1000 simulation runs for each T we also plot the results for the estimator from Kan et al. (2022) ('KWZ') for different values of γ. Both plots show increasing mean out-of-sample Sharpe ratios for all portfolio weight estimators (except for the shrinkage minimum variance portfolio in the left plot) which is to be expected because asset returns are simulated iid over time and population parameters can be estimated with increasing precision. An important insight from Figure 2.2 is that the mean out-of-sample Sharpe ratios of our portfolio strategies ('MaxSR-Approx' and 'MaxSR-Boot') are always higher than the Sharpe ratios of the sample tangency portfolio. The curves approach each

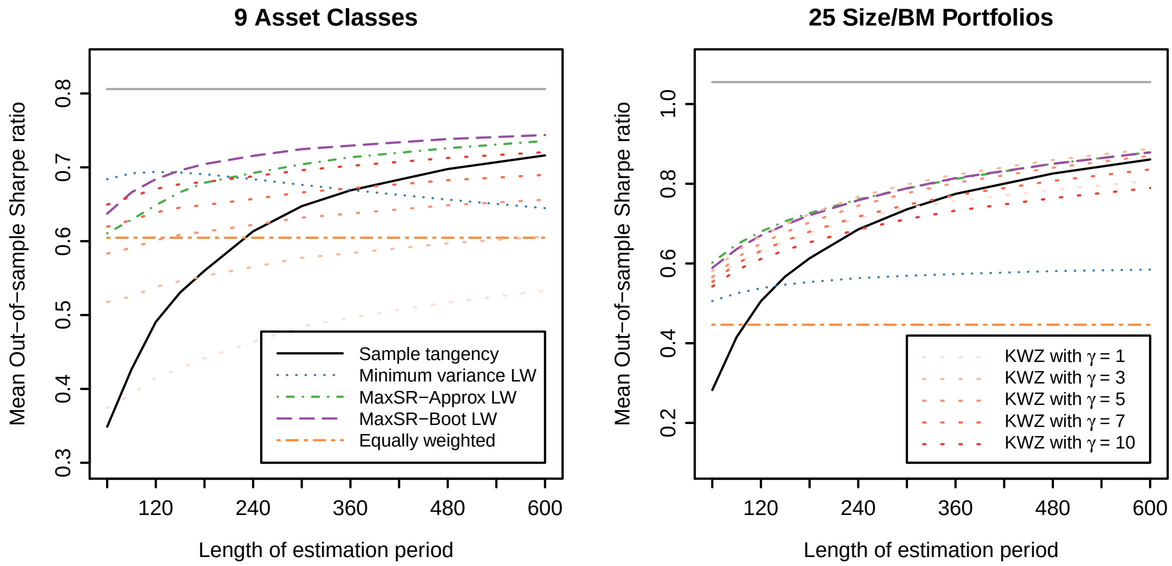

Figure 2.2: Annualized mean out-of-sample Sharpe ratios based on 1000 simulation runs for increasing lengths of estimation periods T and different portfolio strategies. Portfolio strategies are displayed with different colors and line types. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 based on (2.10) and (2.4), respectively, using $\widehat{c}_{\mathrm{PML}}$ from (2.16). 'KWZ' is the estimator of Kan et al. (2022) for different choices of γ. The uppermost horizontal line represents the maximum attainable Sharpe ratio when the population parameters (expected returns and covariance matrix) are known. Population parameters are based on data of 9 Asset Classes (left) and 25 Size/BM portfolios (right), cf. Table 2.1.

other for larger T, but even for $T = 600$, i.e., the equivalent of 50 years of monthly return data, the proposed portfolio strategies yield higher Sharpe ratios than the sample tangency portfolio. The plots show that for the portfolio weight estimators of Kan et al. (2022) it is important to choose a suitable γ to achieve a comparable performance. In the left plot the estimator with $\gamma = 10$ is the only one in this comparison which dominates the tangency portfolio in terms of mean out-of-sample Sharpe ratio for the entire range of T between 60 and 600. A smaller value of γ, implying less of a shift towards the minimum variance portfolio, results in substantially lower Sharpe ratios (by margins of 0.13 to 0.11 in the case of $\gamma = 3$). In the right plot $\gamma = 3$ makes for the highest out-of-sample Sharpe ratios dominating the sample tangency portfolio, while $\gamma = 10$ results in the lowest mean out-of-sample Sharpe ratios. The comparison shows that the optimal value of γ maximizing the out-of-sample Sharpe ratio depends on the specific dataset and is therefore ex-ante unknown. Within our approaches an optimal shrinkage parameter is estimated for each dataset effectively solving the ex-ante specification problem for γ. In the following we set the coefficient of relative risk aversion within the expected utility-based framework of Kan et al. (2022) to $\gamma = 3$ which is a typically chosen value in the literature that in many cases yields a good performance.

Simulation results for all 15 parameter constellations from the datasets in Table 2.1 are summarized based on annualized mean out-of-sample Sharpe ratios in Table 2.2. From Table 2.2 we draw the following conclusions. First, because the weights of the sample tangency portfolio are estimated via the maximum likelihood method and constitute a first attempt at estimating the weights of the population tangency portfolio, it is of interest to examine its Sharpe ratio over all 15 parameter constellations. Table 2.2 shows that the sample tangency portfolio attains on none of the 15 datasets the highest mean out-of-sample Sharpe ratio. The sample tangency portfolio is thus dominated by other estimators in terms of out-of-sample performance. Comparing it with the naive equally weighted portfolio, it obtains a higher Sharpe ratio on 6 datasets while on 9 datasets it attains a lower Sharpe ratio. As the equally weighted portfolio ignores all information in estimated mean returns, variances and correlations, a middle ground would be to use only information from the estimated covariance matrix. The estimated minimum variance portfolio ('MinVar') attains a higher Sharpe ratio on 10 datasets compared to the sample tangency and on 9 datasets compared to the equally weighted portfolio. Hence, balancing information contained in the sample tangency and the minimum variance portfolios may be beneficial, which is what the proposed methods from Section 2.2 aim to do. The comparison of the approximative ('MaxSR-Approx') and the bootstrap-based ('MaxSR-Boot') estimators to the sample tangency and the minimum variance portfolios shows that the former outperform the two latter portfolios on at least 13 of the 15 datasets. In this comparison, the sample minimum variance portfolio attains the highest Sharpe ratio in up to two datasets by only a small margin. Because our aim is to find portfolios which offer consistently high expected out-of-sample Sharpe ratios, we conclude that the proposed methods overall provide for meaningful improvements in out-of-sample Sharpe ratio, by 34 and 22 percent on average, over the sample tangency and the minimum variance portfolios.

Second, as Ledoit and Wolf (2004a) find, out-of-sample performance can be improved by employing shrinkage estimators for the covariance matrix. Table 2.2 shows that the mean out-of-sample Sharpe ratio of the minimum variance portfolio using the shrinkage estimator of Ledoit and Wolf (2004b) ('MinVar-LW') is always higher than or equal to the Sharpe ratio of the minimum variance portfolio using the maximum likelihood estimator of the covariance matrix ('MinVar'). The improvement is generally larger, the larger the number of assets is. The shrinkage minimum variance portfolio attains a higher Sharpe ratio than the equally weighted portfolio on 11 of the 15 datasets. Using a shrinkage estimator of the covariance matrix may therefore improve the estimators from Section 2.2. Table 2.2 shows that the Sharpe ratios of

Table 2.2: Results of the simulation study: Annualized mean out-of-sample Sharpe ratios for 13 portfolio strategies and 15 parameter constellations.

Portfolio strategy	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Optimal	0.49	0.64	0.81	0.63	0.92	1.21	0.56	0.74	0.75	1.06	1.22	0.83	1.65	1.17	1.87
Equal Weights	0.27	0.42	0.60	0.48	0.38	0.39	0.47	0.43	0.49	0.45	0.47	0.46	0.50	0.68	0.56
Sample Tangency	0.04	0.42	0.48	0.33	0.60	0.77	0.25	0.39	0.30	0.49	0.64	0.30	0.94	0.40	0.31
MinVar	0.04	0.43	0.53	0.56	0.68	0.44	0.41	0.45	0.41	0.54	0.62	0.53	0.71	0.60	0.37
MinVar-LW	0.07	0.43	0.69	0.57	0.69	0.47	0.42	0.45	0.42	0.54	0.64	0.57	0.76	0.68	0.67
KWZ-3-LW	0.30	0.52	0.54	0.51	0.75	0.94	0.43	0.54	0.47	0.67	0.79	0.54	1.12	0.71	0.74
MaxSR-Approx c_min	0.28	0.50	0.60	0.55	0.73	0.86	0.43	0.52	0.47	0.64	0.76	0.55	0.96	0.64	0.45
MaxSR-Approx c_pen	0.29	0.49	0.60	0.55	0.73	0.86	0.43	0.52	0.47	0.64	0.76	0.55	0.96	0.64	0.45
MaxSR-Boot c_min	0.29	0.51	0.58	0.52	0.72	0.89	0.42	0.52	0.45	0.64	0.77	0.52	1.02	0.63	0.42
MaxSR-Boot c_pen	0.29	0.51	0.59	0.52	0.72	0.89	0.41	0.52	0.45	0.64	0.77	0.52	1.02	0.63	0.42
MaxSR-Approx-LW c_min	0.30	0.50	0.68	0.55	0.75	0.92	0.44	0.54	0.48	0.67	0.80	0.58	1.08	0.73	0.85
MaxSR-Approx-LW c_pen	0.31	0.50	0.65	0.55	0.75	0.92	0.43	0.54	0.48	0.67	0.80	0.58	1.09	0.73	0.86
MaxSR-Boot-LW c_min	0.30	0.52	0.67	0.53	0.75	0.90	0.44	0.55	0.47	0.67	0.80	0.55	1.11	0.71	0.84
MaxSR-Boot-LW c_pen	0.30	0.51	0.68	0.53	0.75	0.90	0.43	0.55	0.47	0.67	0.80	0.55	1.12	0.71	0.85

Note: Each column refers to a set of parameter values which constitute the population parameter constellation in the simulation and are calibrated on the full time period of the corresponding dataset with number 1–15 in Table 2.1. 'Optimal' denotes the maximum Sharpe Ratio attainable when population parameters are known. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MinVar' denotes the sample minimum variance portfolio. 'KWZ-3-LW' is the estimator from Kan et al. (2022) with $\gamma = 3$. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 from (2.10) and (2.4), where 'c_min' refers to usage of the estimator $\widehat{c}_{\text{min}}$ in (2.14) and 'c_pen' to $\widehat{c}_{\text{PML}}$ in (2.16). The length of the estimation period for each portfolio strategy is $T = 120$. In each column the highest Sharpe ratio is underlined.

our proposed portfolio strategies with shrinkage of the covariance matrix ('MaxSR-Approx-LW' and 'MaxSR-Boot-LW') are always higher than or equal to the Sharpe ratios of our portfolio strategies based on the sample covariance matrix ('MaxSR-Approx' and 'MaxSR-Boot'). The former outperform the sample tangency portfolio on all of the 15 datasets and deliver higher Sharpe ratios compared to the shrinkage minimum variance and equally weighted portfolios on at least 12 datasets. On four datasets the 'MinVar-LW' strategy or the equally weighted portfolio attain the highest Sharpe ratios. In these four cases the Sharpe ratios of the 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' strategies are lower by 0.04 at the maximum. Overall, our proposed portfolio strategies show superior out-of-sample performance compared to the minimum variance and the equally weighted portfolios. Mean out-of-sample Sharpe ratios improve by 22 and 39 percent on average, respectively.

Third, we compare our portfolio strategies to the portfolio weight estimator of Kan et al. (2022) with $\gamma = 3$ ('KWZ-3-LW'). The Sharpe ratios are similar on 13 of the 15 datasets, differing by 0.04 at a maximum. This shows that $\gamma = 3$ often is a good choice for the coefficient of relative risk aversion when the objective is to maximize the out-of-sample Sharpe ratio. It also shows that the performance of our proposed estimators is just as good in these cases, and sometimes slightly better. On datasets #3 and #15 our approximative and bootstrap-based estimators attain higher Sharpe ratios by margins of up to 0.14 and 0.12 compared to 'KWZ-3-LW'. This indicates that our estimators are able to outperform 'KWZ-3-LW' in cases where the choice of $\gamma = 3$ is suboptimal. Among our proposed portfolio strategies we compare the approximative and the bootstrap-based estimators 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' which either employ the estimator $\widehat{c}_{\text{min}}$ from (2.14) ('c_min') or $\widehat{c}_{\text{PML}}$ from (2.16) ('c_pen'). On all 15 datasets the differences in Sharpe ratios are small, i.e., below 0.04. On some datasets the approximative estimator is superior, on others the bootstrap-based estimator, exhibiting no observable systematic. We thus conclude that the performance of the approximative and the bootstrap-based estimators in terms of out-of-sample Sharpe ratio is comparable.

Fourth, Table 2.2 reports the maximum attainable Sharpe ratio for each parameter constellation when population parameters are known. Naturally, the out-of-sample Sharpe ratios are lower for portfolio strategies that need to estimate portfolio weights based on a finite sample. The Sharpe ratios of the 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' strategies are generally closer to the maximum Sharpe ratios compared to the sample tangency, the minimum variance or the equally weighted portfolios. The difference between mean out-of-sample Sharpe ratios and the optimal Sharpe ratio increases with the number of assets. For 10 assets or less the difference in Sharpe ratios is mostly below 0.2 while for more than 25 assets it is mostly larger than 0.4. Thus, parameter estimation risks increase for larger N/T which cannot be fully compensated by employing more sophisticated methods.

Running the simulations for $T = 60$ and $T = 240$ we note that, as expected, higher mean out-of-sample Sharpe ratios are obtained the longer the estimation period, i.e., the larger T. The overall conclusions from the results for $T = 60$ and $T = 240$ are qualitatively similar to the conclusions drawn from Table 2.2.[5] In addition to this, we stress that for $T = 60$ 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' attain higher Sharpe ratios compared to the minimum variance or equally weighted portfolios on 9 of 14 datasets (dataset #15 is left out because here $N > T$) with an average improvement of 11 and 24 percent. For $T = 240$ outperformance is achieved on 14 of 15 datasets (on the one remaining dataset the Sharpe ratios are lower by merely 0.01) with an average improvement of 33 and 55 percent, respectively. The results show that our proposed portfolio strategies strike an advantageous balance between the minimum variance and the tangency portfolios and in most cases achieve higher Sharpe ratios than both, thus generally improving out-of-sample portfolio performance. The performance of the minimum variance and the equally weighted portfolios strongly depend on the actual parameter constellation. While in the left plot in Figure 2.2 the equally weighted portfolio attains higher Sharpe ratios than the sample tangency portfolio for $T < 240$ on dataset #3, it only outperforms the sample tangency portfolio for $T < 120$ in the right plot on dataset #12.

Our proposed estimators also provide an advantage on other datasets over the estimator of Kan et al. (2022) with a fixed $\gamma = 3$ which cannot be observed in Table 2.2 because the table merely focuses on $T = 120$. As Table 2.2 indicates, our estimators attain similar Sharpe ratios when $\gamma = 3$ is a suitable choice and higher Sharpe ratios elsewise. In essence, using the proposed methods we are able to estimate appropriate γ coefficients for each specific dataset. Figure 2.3 depicts mean out-of-sample Sharpe ratios for datasets #13 and #14 for increasing lengths of estimation periods T. Here, $\gamma = 3$ within the estimator of Kan et al. (2022) represents a suitable choice in order to obtain high Sharpe ratios. For $T = 120$ the Sharpe ratio of our bootstrap-based estimator roughly equals the Sharpe ratio of 'KWZ-LW-3'. For larger T, however, our estimators attain higher Sharpe ratios, e.g., by margins of 0.05 when $T = 600$. We observe slight advantages of the bootstrap-based estimator 'MaxSR-Boot-LW' in Figures 2.2 and 2.3 but the two proposed estimators approach each other for larger T and dominate the sample tangency portfolio over the entire range of T. In this context it would be interesting to further improve the calibration of the shrinkage parameter by using longer samples. In the following empirical analysis we nevertheless limit ourselves to an estimation period with $T = 120$ which is the length of estimation period that is typically considered in the literature.

[5] The results are available upon request but are not shown due to the conservation of space.

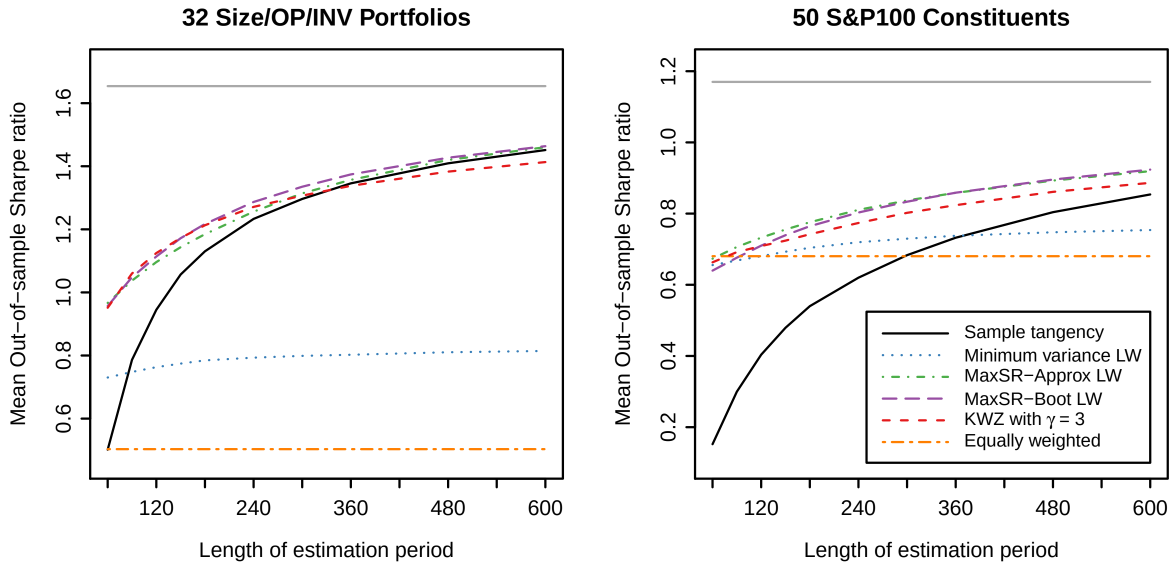

Figure 2.3: Annualized mean out-of-sample Sharpe ratios based on 1000 simulation runs for increasing lengths of estimation periods T and different portfolio strategies. Portfolio strategies are displayed with different colors and line types. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 based on (2.10) and (2.4), respectively, using $\widehat{c}_{\text{PML}}$ from (2.16). 'KWZ' is the estimator of Kan et al. (2022) with $\gamma = 3$. The uppermost horizontal line represents the maximum attainable Sharpe ratio when the population parameters (expected returns and covariance matrix) are known. Population parameters are based on datasets of 32 Size/OP/INV portfolios (left) and 50 constituents of the S&P100 (right), cf. Table 2.1.

2.4 Empirical results

We demonstrate the performance of our portfolio weight estimators from Section 2.2 in a comprehensive empirical analysis. Because transaction costs that occur due to random fluctuations of estimated portfolio weights reduce net returns substantially, we explicitly take accruing transaction costs into account within a second step of the estimation process. After focusing on the performance of portfolios that contain only risky assets, we study the performance of our portfolio allocation method between risky and risk-free assets.

In the analysis we follow a rolling window approach. At each point in time t we use the previous T monthly returns on N risky assets to estimate portfolio weights for the next period. We consider estimation windows of $T = 60$, 120, 240, but for the conservation of space we only report the results for $T = 120$ in the following. In order to estimate portfolio weights, the maximum likelihood estimator for the covariance matrix Σ as well as the shrinkage estimator of Ledoit and Wolf (2004b) are considered as $\widehat{\Sigma}$. When introducing a risk-free asset, we estimate the optimal scaling factor for portfolio weights $\widehat{\lambda^*}$ via a 5-fold cross-validation which is repeated 50 times. In the cross-validation runs we do not re-estimate our optimal shrinkage parameter γ^* because of run time restrictions. The empirical results show that this restriction does not harm the performance of the proposed methods. $\widehat{\lambda^*}$ is estimated as described in Section 2.2.3 irrespective of the subsequent consideration of transaction costs. We set the target level for expected out-of-sample portfolio volatility σ in (2.11) to 0.05 per annum.

Implementing the rolling window approach yields portfolio weights w_t for the N risky assets at each point in time t for each considered portfolio strategy. The resulting realized out-of-sample portfolio excess return for the next period is $w_t' R_{t+1}$, where R_{t+1} denotes the vector of excess returns on the N risky assets from time t to $t+1$. The out-of-sample portfolio excess return net of proportional transaction costs of κ is given by

$$\left(1 + R_{f,t+1} + w_t' R_{t+1}\right) \left(1 - \kappa \sum_{i=1}^{N} |w_{t+1,i} - w_{t^+,i}|\right) - 1 - R_{f,t+1},$$

where $R_{f,t+1}$ is the return on the risk-free asset from time t to $t+1$, and

$$w_{t^+,i} = w_{t,i} \frac{1 + R_{f,t+1} + R_{t+1,i}}{1 + R_{f,t+1} + w_t' R_{t+1}}$$

is the portfolio weight for risky asset i at time $t+1$ immediately before rebalancing. In the following we consider proportional transaction costs in the amount of $\kappa = 50$ bps. 50 bps is a common value in the literature, cf., e.g., DeMiguel et al. (2009), Kirby and Ostdiek (2012), and Olivares-Nadal and DeMiguel (2018). When including transaction costs in the estimation process as in (2.12) and (2.13), we set the vector of prevailing portfolio weights w_0 to w_{t^+} at each point in time t, and κ in (2.12) and (2.13) is equivalent to $\kappa = 50$ bps.

From the resulting out-of-sample portfolio returns as well as from portfolio net returns we calculate means, standard deviations and Sharpe ratios for each of the considered portfolio strategies. The portfolio strategies are primarily compared based on their realized out-of-sample (net) Sharpe ratios which we present in Table 2.3. From Panel A in Table 2.3 we draw the following conclusions. First, as DeMiguel et al. (2009) document, the Sharpe ratios of the sample tangency portfolio are substantially lower than the Sharpe ratios of the equally weighted portfolio. Across the 15 datasets its Sharpe ratios are lower on all datasets except #13, in many of the datasets its Sharpe ratios are negative indicating a negative mean return over the out-of-sample period. There is a large discrepancy between the Sharpe ratio of the out-of-sample optimal portfolio, which is based on the set of out-of-sample returns and thus uses information from the 'future', and the estimated tangency portfolio. The sample tangency portfolio produces extreme portfolio weights and therefore high portfolio volatility. We find that the sample minimum variance portfolio outperforms the sample tangency portfolio on 14 and the equally weighted portfolio on 7 of the 15 datasets. The portfolio strategies 'MaxSR-Approx' and 'MaxSR-Boot' outperform the sample tangency portfolio on all, the equally weighted portfolio on 9 and the minimum variance portfolio on up to 10 of the 15 datasets.

Second, usage of the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) improves the empirical out-of-sample performance of the considered portfolio strategies. The shrinkage minimum variance portfolio ('MinVar-LW') delivers superior Sharpe ratios on all of the 15 datasets compared to the sample tangency portfolio and on 12 datasets compared to the equally weighted portfolio with an average improvement of 15 percent. We therefore conclude that there is value in using sample-based information for portfolio optimization, e.g., when ignoring expected return estimates as other literature has found. By making use of expected return estimates our proposed strategies 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' outperform the 'MinVar-LW' strategy on 8 and the equally weighted portfolio on up to 12 of the 15 datasets. Out-of-sample Sharpe ratios improve by 15 percent and 32 percent on average compared with the minimum variance and equally weighted portfolios, respectively. When compared to the estimator of Kan et al. (2022) with $\gamma = 3$ ('KWZ-3-LW') our strategies attain higher out-of-sample Sharpe ratios on up to 9 datasets. Higher Sharpe ratios by margins of up to 0.28, 0.08, 0.15 and 0.08 are attained on datasets #3, #7, #9 and #15 for the approximative ('MaxSR-Approx-LW') or bootstrap-based ('MaxSR-Boot-LW') estimators, respectively. On the remaining datasets our strategies often yield comparable Sharpe ratios, as observed in the simulation study in Section 2.3. This shows empirically that $\gamma = 3$ often is an appropriate choice for expected utility-based approaches if the objective is to maximize the out-of-sample Sharpe ratio. Overall, our estimators which directly aim to maximize out-of-sample Sharpe ratios are able to achieve favorable Sharpe ratios in cases where $\gamma = 3$ represents a suboptimal choice and perform just as good in all other cases.

Third, we test for the statistical significance of the differences in Sharpe ratios between our 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' strategies and the shrinkage minimum variance or the equally weighted portfolio strategies using the asymptotic test of Ledoit and Wolf (2008). A statistically significant difference of Sharpe ratios on the 5% level against the minimum variance portfolio is highlighted in italic and against the equally weighted portfolio in bold in Panel A of Table 2.3. Our approximative estimator significantly outperforms the minimum variance portfolio on up to 6 and the equally weighted portfolio on 7 datasets. Our bootstrap-based estimator significantly outperforms the minimum variance portfolio on up to 5 and the equally weighted portfolio on 4 datasets. A statistically significant outperformance is often achieved on datasets which on the one hand, as Kirby and Ostdiek (2012) note, display a considerable

Table 2.3: Results of the empirical analysis: Annualized out-of-sample (net) Sharpe ratios for 13 portfolio strategies and 15 datasets.

Portfolio strategy	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Out-of-Sample Optimal	0.61	0.45	0.98	0.70	1.01	1.28	0.75	0.83	0.93	1.26	1.36	0.89	1.71	1.36	1.93
Panel A: Annualized Sharpe Ratios															
Equal Weights	0.32	0.36	0.62	0.57	0.45	0.44	0.61	0.51	0.54	0.54	0.52	0.54	0.55	0.82	0.51
Sample Tangency	−0.08	−0.02	0.40	0.11	0.18	−0.02	0.19	−0.03	−0.16	0.14	0.01	−0.09	0.76	−0.11	−0.04
MinVar	0.05	0.41	0.52	0.56	0.64	0.27	0.65	0.50	0.54	0.74	0.66	0.51	0.69	0.51	0.47
MinVar-LW	0.09	0.39	0.85	0.61	0.64	0.39	0.66	0.52	0.60	0.75	0.72	0.61	0.83	0.64	0.76
KWZ-3-LW	*0.53*	0.20	0.54	0.50	***0.93***	***1.05***	0.59	0.66	*0.35*	***1.00***	0.80	0.54	***1.16***	0.58	***0.83***
MaxSR-Approx c_min	0.59	0.10	0.55	0.53	0.87	0.98	0.65	0.56	0.48	0.87	0.73	0.52	0.95	0.49	0.56
MaxSR-Approx c_pen	0.56	0.13	0.42	0.54	0.87	1.02	0.65	0.38	0.48	0.88	0.73	0.52	0.95	0.49	0.56
MaxSR-Boot c_min	0.57	0.16	0.54	0.48	0.87	1.07	0.61	0.56	0.36	0.83	0.62	0.43	0.95	0.49	0.56
MaxSR-Boot c_pen	0.54	0.18	0.54	0.48	0.85	1.11	0.63	0.43	0.37	0.79	0.61	0.42	0.97	0.49	0.56
MaxSR-Approx-LW c_min	***0.60***	0.00	0.82	0.57	***0.88***	***1.09***	0.65	0.62	0.50	***0.96***	**0.82**	0.59	***1.16***	0.56	***0.90***
MaxSR-Approx-LW c_pen	***0.57***	0.05	0.68	0.58	**0.86**	***1.08***	0.67	0.51	0.50	***0.97***	**0.82**	0.59	***1.16***	0.56	***0.91***
MaxSR-Boot-LW c_min	*0.56*	0.17	0.71	0.54	***0.89***	***1.08***	0.64	0.65	0.42	***0.93***	0.74	0.51	***1.17***	0.51	0.76
MaxSR-Boot-LW c_pen	*0.55*	0.16	0.74	0.53	***0.87***	***1.12***	0.65	0.58	0.45	**0.89**	0.73	0.51	***1.18***	0.53	0.79

Table 2.3 (continued)

Portfolio strategy	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Panel B: Annualized Net Sharpe Ratios															
Equal Weights	0.30	0.35	0.60	0.56	0.45	0.43	0.61	0.50	0.53	0.53	0.51	0.53	0.54	0.80	0.51
Sample Tangency	−0.42	−0.27	−0.92	−0.12	−0.06	−0.15	−0.16	−0.32	−0.21	−0.22	−0.28	−0.24	−0.06	−0.43	−0.12
MinVar	0.03	0.38	−0.44	0.48	0.53	0.14	0.53	0.39	0.44	0.40	0.43	0.30	0.32	0.29	−1.20
MinVar-LW	0.07	0.37	0.78	0.55	0.58	0.32	0.58	0.46	0.52	0.59	0.57	0.47	0.63	0.52	0.29
KWZ-3-LW	0.46	0.13	0.36	0.34	0.72	0.81	0.40	0.44	0.19	0.59	0.45	0.26	0.71	0.46	0.34
MaxSR-Approx c_min	0.48	−0.02	−0.60	0.38	0.63	0.58	0.48	0.37	0.34	0.48	0.46	0.27	0.52	0.26	−1.10
MaxSR-Approx c_pen	0.46	0.03	−0.57	0.39	0.62	0.60	0.48	0.15	0.35	0.48	0.46	0.27	0.53	0.26	−1.10
MaxSR-Boot c_min	0.47	0.07	−0.58	0.33	0.58	0.66	0.38	0.28	0.16	0.26	0.22	0.11	0.32	0.26	−1.10
MaxSR-Boot c_pen	0.45	0.08	−0.58	0.32	0.55	0.65	0.37	0.15	0.17	0.20	0.20	0.10	0.35	0.26	−1.10
MaxSR-Approx-LW c_min	0.50	−0.10	0.74	0.45	0.67	0.85	0.50	0.44	0.39	0.64	0.59	0.41	0.82	0.43	0.36
MaxSR-Approx-LW c_pen	0.47	−0.06	0.54	0.46	0.64	0.78	0.51	0.31	0.39	0.65	0.59	0.41	0.82	0.43	0.37
MaxSR-Boot-LW c_min	0.47	0.10	0.62	0.42	0.69	0.88	0.49	0.46	0.27	0.57	0.41	0.28	0.73	0.38	0.00
MaxSR-Boot-LW c_pen	0.47	0.07	0.66	0.41	0.67	0.90	0.49	0.37	0.30	0.53	0.40	0.28	0.76	0.41	0.03

Note: Each column refers to a dataset with number 1–15 in Table 2.1. 'Out-of-Sample Optimal' denotes the Sharpe ratio of the portfolio whose weights are estimated from the full set of out-of-sample asset returns. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MinVar' denotes the sample minimum variance portfolio. 'KWZ-3-LW' is the estimator of Kan et al. (2022) with $\gamma = 3$. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 from (2.10) and (2.4), where 'c_min' refers to usage of the estimator $\widehat{c}_{\min}$ in (2.14) and 'c_pen' to $\widehat{c}_{\text{PML}}$ in (2.16). The results are based on an estimation window of $T = 120$ months. In Panel A we employ the asymptotic test with HAC standard errors of Ledoit and Wolf (2008) to test for different Sharpe ratios of 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' compared to 'MinVar-LW' and the equally weighted portfolio. Statistically significant differences on the 5% level against 'MinVar-LW' are displayed in italic and against the equally weighted portfolio in bold. Panel B reports annualized net Sharpe ratios after incorporating transaction costs in the amount of 50 bps. In each column the highest Sharpe ratio is underlined.

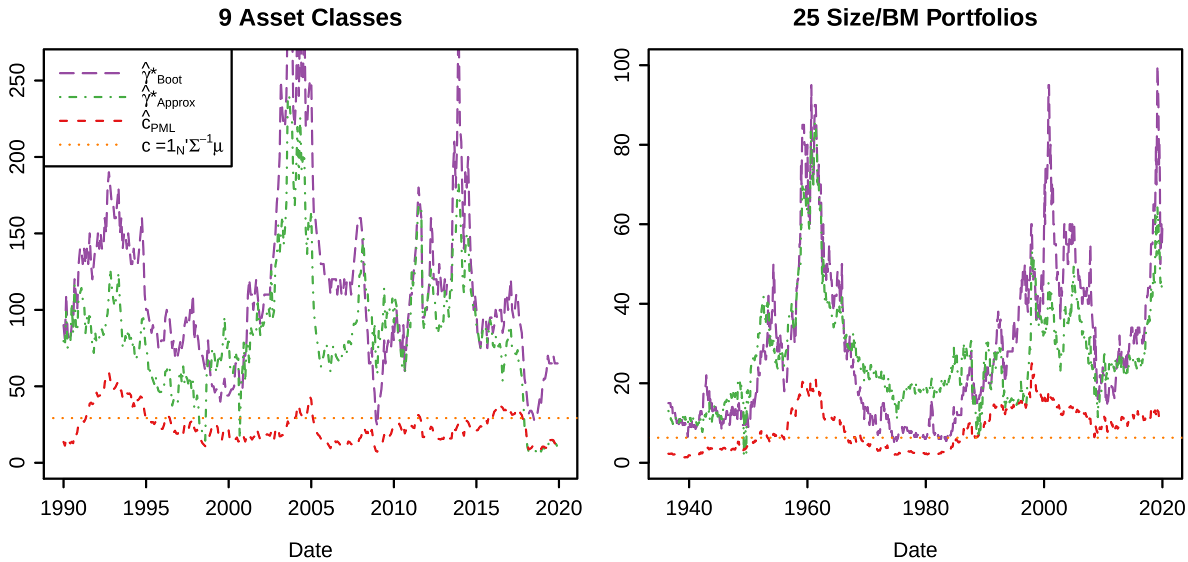

Figure 2.4: Estimated $\widehat{c}_{\text{PML}}$ from (2.16) and $\widehat{\gamma^*}$ from the proposed approximative, $\widehat{\gamma^*}_{\text{Approx}}$, and bootstrap-based, $\widehat{\gamma^*}_{\text{Boot}}$, estimators from Section 2.2.2 (see (2.10) and (2.4), respectively) which use $\widehat{c}_{\text{PML}}$, displayed over the out-of-sample periods of two datasets (cf. Table 2.1). The estimates are based on an estimation window of $T = 120$ past returns. The population value $c = 1_N'\Sigma^{-1}\mu$, displayed as a horizontal dotted line, is estimated from the full set of out-of-sample returns.

amount of cross-sectional dispersion in the means and variances of returns, and on the other hand possess a long enough return history to detect such differences at the specified confidence level. The 'KWZ-3-LW' strategy displays a similar performance in terms of statistical significance on these datasets, but significantly underperforms the minimum variance portfolio on dataset #9 with a p-value of 0.01. Our proposed estimators show no significant underperformance either compared to the minimum variance or the equally weighted portfolio.

To shed more light on the functioning of our estimators, we depict the estimates of the shrinkage parameter, $\widehat{\gamma^*}$ from Section 2.2.2, over the out-of-sample periods of two datasets in Figure 2.4. The estimates from the approximative (dot-dashed) and bootstrap-based (long-dashed) methods are displayed together with estimates of $c = 1_N'\Sigma^{-1}\mu$, the denominator of the population tangency portfolio in (2.1). The estimates of c are based on our (approximate) penalized maximum likelihood estimator $\widehat{c}_{\text{PML}}$ from (2.16) (short-dashed), and the full set of out-of-sample returns (dotted) which uses 'future' information and should thus be the best estimate of c. Figure 2.4 shows that $\widehat{c}_{\text{PML}}$ fluctuates around c which is to be expected because it only uses information from the past $T = 120$ returns. The level of c in the left plot is notably higher than in the right plot (29.2 vs. 6.3) because $c = 1_N'\Sigma^{-1}\mu = \mu_{\text{MINV}}/\sigma^2_{\text{MINV}}$ and the 9 asset classes contain low volatility assets leading to a lower variance of the minimum variance portfolio, σ^2_{MINV}, while the 25 Size/BM dataset contains only equity portfolios. Because c is the effective risk aversion coefficient maximizing the Sharpe ratio for the true efficient frontier without parameter uncertainty, an estimate of c such as $\widehat{c}_{\text{PML}}$, which is used in our approximative and bootstrap-based procedures, is the first component of our shrinkage parameter estimates $\widehat{\gamma^*}$. The second component is an add-on resulting from the presence of parameter estimation risks. $\widehat{\gamma^*}$ is therefore always greater than $\widehat{c}_{\text{PML}}$, i.e., $\widehat{\gamma^*}/\widehat{c}_{\text{PML}} > 1$. The magnitude of the ratio depends on the trade-off of two factors. The higher the estimation risks from estimated mean returns $\widehat{\mu}$ and the lower the expected improvement over the minimum variance portfolio by shifting towards the tangency portfolio, the higher is $\widehat{\gamma^*}/\widehat{c}_{\text{PML}}$. The magnitude of estimation risks from $\widehat{\mu}$ depends on the ratio N/T, i.e., number of assets relative to length of estimation period, and the dispersion of expected returns μ. Figure 2.4 shows that the ratio $\widehat{\gamma^*}/\widehat{c}_{\text{PML}}$ is larger for the dataset of 9 asset classes than for the 25 Size/BM portfolios (roughly 5 vs. 4) despite the lower number of assets. The reason is that, because volatilities are in line with expected returns for asset classes, the expected improvement over the minimum variance portfolio by shifting towards the estimated tangency portfolio will not be as large as for the 25 Size/BM portfolios who by construction contain a reasonable amount of dispersion in expected returns. Thus, in Panel A of Table 2.3 the minimum variance portfolio performs best for the 9 asset classes (dataset #3) while our estimators MaxSR-Approx-LW' and 'MaxSR-Boot-LW' outperform 'KWZ-3-LW' because in our procedures we estimate c, via $\widehat{c}_{\min}$ from (2.14) or $\widehat{c}_{\text{PML}}$ from (2.16), which is much higher than $\gamma = 3$ on this dataset. Seeing the rather large fluctuations of the parameter estimates of $\widehat{c}_{\text{PML}}$ and $\widehat{\gamma^*}$ in Figure 2.4, a means of further improvement could be to consider longer asset return time series solely for estimation of the optimal shrinkage parameter γ^*.

Panel B of Table 2.3 reports annualized net Sharpe ratios after incorporating transaction costs in the amount of 50 bps. The impact of transaction costs on the out-of-sample Sharpe ratios of the sample tangency portfolio is expectedly large since it produces extreme and unstable portfolio weights as, e.g., previously noted in DeMiguel et al. (2009) and Kirby and Ostdiek (2012). Its poor out-of-sample performance thus deteriorates even further when transaction costs are considered. The 'MinVar-LW' strategy displays a substantially lower reduction in Sharpe ratio due to transaction costs because it avoids estimation risks stemming from expected returns estimates. For datasets containing less than 25 assets, the reduction in Sharpe ratio is less than 0.07 while for larger datasets it exceeds 0.14. For instance, for dataset #15 containing 100 assets the annualized net Sharpe ratio is 0.29 whereby the Sharpe ratio ignoring transaction costs is 0.76 which makes for a large reduction in Sharpe ratio of 0.47. The equally weighted portfolio naturally requires little turnover which is often cited as one of its advantages, e.g., in DeMiguel et al. (2009). Its reduction in Sharpe ratio due to transaction costs is lower than 0.02 on all datasets. The implementation of our 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' strategies results in higher turnover compared to 'MinVar-LW'. Its Sharpe ratios decrease by 0.07 to 0.30 for $N < 25$ and by 0.13 to 0.76 for $N \geq 25$ due to accruing transaction costs. The 'MaxSR-Boot-LW' method often leads to slightly higher turnover compared to 'MaxSR-Approx-LW' with the largest difference on dataset #15 (a by 0.75 lower net Sharpe ratio for 'MaxSR-Boot-LW' and by 0.54 for 'MaxSR-Approx-LW' due to transaction costs). For 'KWZ-3-LW' the reduction in net Sharpe ratio due to transactions costs is mostly higher than for our strategies. For instance, 'MaxSR-Approx-LW c_min' attains higher net Sharpe ratios than 'KWZ-3-LW' on 12 of the 15 datasets. Transaction costs are an important factor in out-of-sample performance when comparing the net Sharpe ratios of 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' with 'MinVar-LW' and the equally weighted portfolio. In terms of net Sharpe ratios 'MaxSR-Approx-LW' outperforms 'MinVar-LW' and the equally weighted portfolio on merely 6 to 7 of the 15 datasets. 'MaxSR-Boot-LW' only compares favorably to 'MinVar-LW' and the equally weighted portfolio on 4 to 6 of the 15 datasets. This observation motivates the consideration of transaction costs within the estimation of portfolio weights which we have done following the approach detailed in Section 2.2.4.

The results are reported in Table 2.4. The table compares the approximative 'MaxSR-Approx-LW' and bootstrap-based 'MaxSR-Boot-LW' strategies that incorporate transaction costs with the 'MinVar-LW' strategy and the equally weighted portfolio based on their out-of-sample net Sharpe ratios. When comparing the results from Table 2.4 with the results from Panel B of Table 2.3, the net Sharpe ratios of the estimators are almost always improved (except for dataset #1) when transaction costs are incorporated into the estimation process. We often achieve gains in out-of-sample net Sharpe ratios by margins of 0.10 and more with an average increase by 0.19, i.e., 40 percent. The table shows that our simple approach as given in (2.12) in Section 2.2.4 works well empirically. We achieve a substantial reduction in turnover for our strategies when incorporating transaction costs. Specifically, Sharpe ratios ignoring transaction costs differ from reported net Sharpe ratios by 0.03 at a maximum for datasets #1 to #14 and by up to 0.07 for dataset #15. Incorporating transaction costs does not harm the performance of the strategies. 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' obtain higher net Sharpe ratios compared with 'MinVar-LW' on up to 13 of the 15 datasets and on up to 12 datasets compared with the equally weighted portfolio providing average improvements of 37 and 29 percent, respectively. Table 2.4 reports in italic the p-values of the asymptotic test for differences in Sharpe ratios of Ledoit and Wolf (2008). The results show that 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' significantly outperform 'MinVar-LW' on 7 to 8 of 15 datasets and the equally weighted portfolio on 5 to 6 datasets on the 5% level. On the 10% level these figures rise to up to 9 and up to 7 of 15 datasets, respectively. On none of the considered datasets neither the shrinkage minimum variance portfolio 'MinVar-LW' nor the equally weighted portfolio achieve a significantly higher net Sharpe ratio on the 5% level compared to our proposed methods.

Table 2.4: Results of the empirical analysis: Annualized out-of-sample net Sharpe ratios for the proposed portfolio strategies incorporating transaction costs as in (2.12) in the amount of $\kappa = 50$ bps.

Portfolio strategy	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Equal Weights	0.30	0.35	0.60	0.56	0.45	0.43	0.61	0.50	<u>0.53</u>	0.53	0.51	<u>0.53</u>	0.54	<u>0.80</u>	0.51
MinVar-LW	0.07	<u>0.37</u>	0.78	0.55	0.58	0.32	0.58	0.46	0.52	0.59	0.57	0.47	0.63	0.52	0.29
MaxSR-Approx-LW c_min	<u>0.38</u>	0.33	<u>0.87</u>	0.57	0.77	0.86	0.62	<u>0.63</u>	0.48	0.85	0.85	0.51	1.12	0.56	0.77
p-value vs. EW	*0.40*	*0.76*	*0.14*	*0.88*	*0.00*	*0.03*	*0.83*	*0.08*	*0.59*	*0.00*	*0.01*	*0.90*	*0.00*	*0.15*	*0.03*
p-value vs. MinVar-LW	*0.00*	*0.62*	*0.12*	*0.63*	*0.02*	*0.00*	*0.44*	*0.02*	*0.49*	*0.00*	*0.00*	*0.33*	*0.00*	*0.53*	*0.00*
MaxSR-Approx-LW c_pen	0.35	0.22	0.83	<u>0.58</u>	<u>0.77</u>	0.89	<u>0.64</u>	0.57	0.48	<u>0.86</u>	0.85	0.51	<u>1.13</u>	0.57	<u>0.78</u>
p-value vs. EW	*0.63*	*0.12*	*0.20*	*0.76*	*0.00*	*0.02*	*0.62*	*0.52*	*0.58*	*0.00*	*0.01*	*0.86*	*0.00*	*0.15*	*0.03*
p-value vs. MinVar-LW	*0.01*	*0.18*	*0.55*	*0.44*	*0.02*	*0.00*	*0.31*	*0.29*	*0.47*	*0.00*	*0.00*	*0.39*	*0.00*	*0.52*	*0.00*
MaxSR-Boot-LW c_min	0.38	0.36	0.81	0.58	0.77	0.91	0.62	0.63	0.38	0.84	0.85	0.50	1.07	0.52	0.73
p-value vs. EW	*0.38*	*0.94*	*0.21*	*0.85*	*0.00*	*0.01*	*0.90*	*0.07*	*0.14*	*0.00*	*0.02*	*0.82*	*0.00*	*0.12*	*0.07*
p-value vs. MinVar-LW	*0.00*	*0.91*	*0.72*	*0.56*	*0.01*	*0.00*	*0.49*	*0.02*	*0.10*	*0.00*	*0.01*	*0.49*	*0.00*	*1.00*	*0.00*
MaxSR-Boot-LW c_pen	0.35	0.25	0.83	0.58	0.77	<u>0.93</u>	0.62	0.60	0.36	0.84	<u>0.85</u>	0.51	1.08	0.53	0.73
p-value vs. EW	*0.63*	*0.20*	*0.19*	*0.82*	*0.00*	*0.01*	*0.81*	*0.28*	*0.10*	*0.00*	*0.01*	*0.83*	*0.00*	*0.13*	*0.08*
p-value vs. MinVar-LW	*0.01*	*0.22*	*0.58*	*0.51*	*0.01*	*0.00*	*0.46*	*0.13*	*0.07*	*0.00*	*0.00*	*0.48*	*0.00*	*0.90*	*0.00*

Note: Each column refers to a dataset with number 1–15 in Table 2.1. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MinVar-LW' denotes the estimated shrinkage-based minimum variance portfolio. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 from (2.10) and (2.4), where 'c_min' refers to usage of the estimator $\widehat{c}_{\text{min}}$ in (2.14) and 'c_pen' to $\widehat{c}_{\text{PML}}$ in (2.16). For each of these portfolio strategies the p-value of the statistical test for different Sharpe ratios of the respective strategy and the minimum variance or the equally weighted portfolio is reported. We use the asymptotic test methodology with HAC standard errors from Ledoit and Wolf (2008). Portfolio weights are estimated from a window of $T = 120$ monthly asset returns. In each column the highest Sharpe ratio is underlined.

The final aim is to extend the proposed portfolio strategies such that they satisfy a prespecified risk constraint. The methodology is detailed in Section 2.2.3. We include transaction costs within portfolio estimation as described in (2.13) in the amount of $\kappa = 50$ bps. When evaluating the resulting portfolio strategies, two key figures are of primary interest to us. On the one hand the portfolio strategies aim to achieve high out-of-sample net Sharpe ratios, on the other hand their annualized out-of-sample net return volatility should not exceed the preset risk constraint of $\sigma = 0.05$ per annum. Table 2.5 presents the empirical results in terms of net Sharpe ratios (Panel A) and volatilities (Panel B). Panel A shows that the proposed portfolio strategies 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' attain higher Sharpe ratios than 'MinVar-LW' on 11 of 15 datasets and on 11 to 12 datasets compared with the equally weighted portfolio. Out-of-sample net Sharpe ratios improve by 41 and 39 percent on average, respectively. A statistically significant outperformance of the shrinkage minimum variance portfolio on the 5% level (indicated in italic) is achieved on up to 8 datasets. A significant outperformance of the equally weighted portfolio on the 5% level (indicated in bold) is achieved on 6 of the 15 datasets. Considering the 10% significance level, a statistically significant outperformance is achieved on up to 9 and on 7 of the 15 datasets, respectively. On none of the 15 datasets are the net Sharpe ratios of our proposed portfolio strategies significantly lower than the net Sharpe ratios of the shrinkage minimum variance or the equally weighted portfolios on the 5% level. We conclude that the proposed portfolio strategies offer an overall favorable out-of-sample performance in terms of net Sharpe ratios. Concerning the volatilities of the considered portfolio strategies reported in Panel B of Table 2.5 we observe random-like variations of the realized out-of-sample volatilities around the prespecified limit of expected out-of-sample volatility of 5% per annum. The risk constraint as formalized in (2.11) turns out to be satisfied empirically. The volatilities of net returns of the 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' strategies exceed the risk constraint by at most 0.7% per annum. The scaled shrinkage minimum variance or the equally weighted portfolios do not display lower variations of annualized out-of-sample volatilities.

Summarizing the empirical results presented in this section, we find that the proposed approximative and bootstrap-based estimators for the Sharpe ratio maximizing portfolio exhibit an overall promising empirical out-of-sample performance. The empirical results confirm the findings from the simulation study reported in Section 2.3. We find that transaction costs of non-negligible amounts substantially diminish out-of-sample net Sharpe ratios of our portfolio strategies. The proposed estimators attain higher Sharpe ratios than the estimated minimum variance or the equally weighted portfolios on 8 to 12 of the 15 considered datasets with an average improvement of 15 and 32 percent, respectively, when transaction costs are ignored. Higher net Sharpe ratios are obtained on only 4 to 7 datasets. When considering transaction costs within the estimation process as described in Section 2.2.4, net Sharpe ratios increase

Table 2.5: Results of the empirical analysis: Annualized out-of-sample net Sharpe ratios and volatilities for the proposed portfolio strategies with risk-free asset incorporating transaction costs as in (2.13) in the amount of $\kappa = 50$ bps.

Portfolio strategy	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Panel A: Annualized Net Sharpe Ratios															
Equal Weights	0.25	0.30	0.54	0.56	0.45	0.40	0.56	0.50	0.52	0.54	0.47	0.53	0.50	0.75	0.48
MinVar-LW	−0.02	0.31	0.74	0.59	0.59	0.30	0.58	0.48	0.55	0.61	0.55	0.52	0.62	0.53	0.29
MaxSR-Approx-LW c_min	*0.39*	0.05	0.82	0.59	***0.94***	***0.90***	0.56	*0.64*	0.47	***0.95***	***0.84***	0.55	***1.06***	0.56	***0.79***
MaxSR-Approx-LW c_pen	*0.36*	0.01	0.79	0.58	***0.94***	***0.94***	0.57	0.63	0.47	***0.95***	***0.85***	0.55	***1.05***	0.56	***0.80***
MaxSR-Boot-LW c_min	*0.37*	0.15	0.74	0.59	***0.99***	***1.01***	0.58	*0.66*	0.41	***0.92***	***0.85***	0.51	***1.11***	0.55	***0.87***
MaxSR-Boot-LW c_pen	*0.38*	0.08	0.77	0.58	***0.98***	***1.03***	0.58	*0.68*	0.39	***0.92***	***0.84***	0.51	***1.12***	0.56	***0.86***
Panel B: Annualized Volatilities of Net Returns (in percent)															
Equal Weights	5.0	5.3	5.4	4.9	5.0	5.1	5.0	4.9	5.4	4.9	5.2	5.0	5.2	5.1	5.3
MinVar-LW	5.1	5.4	4.9	5.3	5.0	5.4	5.2	5.0	5.7	5.2	5.5	5.3	5.6	5.3	5.5
MaxSR-Approx-LW c_min	5.1	4.8	4.8	5.1	4.9	5.3	5.4	4.9	5.7	5.3	5.4	5.1	5.7	5.2	5.3
MaxSR-Approx-LW c_pen	5.1	4.5	4.9	5.1	4.8	5.4	5.3	4.9	5.7	5.3	5.4	5.1	5.7	5.2	5.3
MaxSR-Boot-LW c_min	5.0	5.1	4.8	5.1	4.8	5.2	5.2	4.8	5.4	5.2	5.0	4.9	5.4	5.3	4.8
MaxSR-Boot-LW c_pen	5.0	4.8	4.8	5.1	4.8	5.2	5.2	4.8	5.5	5.2	5.1	4.9	5.4	5.3	4.8

Note: Each column refers to a dataset with number 1–15 in Table 2.1. 'LW' indicates that the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004b) is used. 'MinVar-LW' denotes the estimated shrinkage-based minimum variance portfolio. 'MaxSR-Approx' and 'MaxSR-Boot' are the proposed approximative and bootstrap-based estimators from Section 2.2.2 from (2.10) and (2.4), where 'c_min' refers to usage of the estimator $\widehat{c}_{\text{min}}$ in (2.14) and 'c_pen' to $\widehat{c}_{\text{PML}}$ in (2.16). Portfolio weights are estimated from a window of $T = 120$ monthly asset returns. In Panel A we employ the asymptotic test with HAC standard errors of Ledoit and Wolf (2008) to test for different Sharpe ratios of 'MaxSR-Approx-LW' and 'MaxSR-Boot-LW' compared to 'MinVar-LW' and the equally weighted portfolio. Statistically significant differences on the 5% level against 'MinVar-LW' are displayed in italic and against the equally weighted portfolio in bold. In each column the highest Sharpe ratio is underlined.

considerably and the portfolio strategies are able to outperform the minimum variance or equally weighted portfolios on up to 12 of the 15 datasets with an average improvement of 37 and 29 percent, respectively. Beyond that, we find that a prespecified portfolio risk limit can be satisfied without sacrificing out-of-sample performance. While holding expected portfolio volatility fixed at 5% per annum, the proposed estimators outperform the minimum variance portfolio on 11 datasets (with 8 being significant on the 5% level) and the equally weighted portfolio on 11 to 12 datasets (with 6 being significant on the 5% level) in terms of out-of-sample net Sharpe ratios with an average improvement of 41 and 39 percent, respectively. We find that portfolio optimization strategies that explicitly take parameter estimation risks into account are able to add value to and outperform simplified or non-optimization-based portfolio diversification approaches.

2.5 Conclusion

In this paper we consider a new calibration criterion for shrinkage portfolio weight estimators dealing with parameter estimation risks. Because the Sharpe ratio is an important performance measure in academia and practice, we aim to directly maximize the expected out-of-sample Sharpe ratio of estimated portfolio weights. We apply an approximative and an improved bootstrap-based method to estimate the optimal shrinkage parameter. The shrinkage parameter is the risk aversion coefficient tracing out the in-sample efficient frontier. We provide an interpretation of the optimal shrinkage parameter as an implicit effective risk aversion coefficient consisting of the value that maximizes the Sharpe ratio for the true efficient frontier without parameter uncertainty and an add-on resulting from the presence of parameter estimation risks. Thus, if the objective is to maximize the out-of-sample Sharpe ratio, the proposed methods avoid the need of deciding on an effective risk aversion coefficient, which is required when shrinkage portfolios are based on an expected utility criterion as in Kan and Zhou (2007) and Kan et al. (2022). The favorable performance of the portfolio weight estimators in terms of out-of-sample Sharpe ratio is demonstrated based on a simulation study and empirical analysis with 15 datasets. We find that our portfolio strategies are able to add value and, overall, outperform the minimum variance and equally weighted portfolios.

When implementing portfolio strategies in practice, an important concern is the amount of produced portfolio turnover which may lead to significant transaction costs diminishing realized net Sharpe ratios. Our portfolio strategies taking estimation errors into account still produce significant amounts of turnover. In the empirical analysis we show that considering transaction costs ex-ante within the estimation process substantially improves the out-of-sample performance of portfolio strategies developed under parameter uncertainty. Exhibiting a two-fund separation for the case with a risk-free asset, we explore an extension of the optimization problem that includes a constraint for expected out-of-sample portfolio volatility. The required scaling factor for portfolio weights of risky assets is determined via a 5-fold cross-validation. We show empirically that out-of-sample portfolio volatilities are consistent with the prespecified volatility target. Favorable out-of-sample performance of the strategies in terms of realized net Sharpe ratios remains intact. In summary, the proposed portfolio weight estimators significantly outperform the minimum variance portfolio using the shrinkage estimator of Ledoit and Wolf (2004b) and the equally weighted portfolio on a 5% level on 8 and 6 of the 15 considered datasets, respectively. Out-of-sample net Sharpe ratios improve by 41 and 39 percent on average.

Observing the variation of our estimates of the optimal shrinkage parameter over time, the proposed methods could further benefit from considering longer return time series for parameter estimation.

Appendix 2.A

2.A.1 Estimating the shrinkage intensity for the mean vector

In order to estimate the optimal shrinkage intensity α^* from Section 2.2.2, (2.6) shows that the plug-in estimator of $(\mu - \bar{\mu}\, 1_N)'\,(\mu - \bar{\mu}\, 1_N)$ is biased upwards. Its unbiased version is inadmissible because it can take negative values. Thus, we aim to adjust the unbiased estimator such that it is positive and approaches the unbiased estimator for large values. To achieve this, we define

$$X = \frac{(\widehat{\mu} - \widetilde{\mu}\, 1_N)'\,(\widehat{\mu} - \widetilde{\mu}\, 1_N)}{\left(\mathrm{trace}(\Sigma)/N - 1_N' \Sigma\, 1_N / N^2\right)/T}$$

which implies that $X - N$ is unbiased. We then construct an estimator for α^* of the form

$$\widehat{\alpha^*} = 1 - \sqrt{\frac{\widetilde{X}}{X}},$$

where

$$\widetilde{X} = X - N + \frac{2X^{N/2}\,\exp\{-X/2\}}{\int_0^X t^{N/2-1}\,\exp\{-t/2\}\,\mathrm{d}t}.$$

The right-hand term in $\widetilde{X}$ is based on an estimator suggested in Kubokawa et al. (1993), where the integral in the denominator is proportional to the χ^2-distribution function with parameter N.[6] Σ in X is estimated via $\widehat{\Sigma}$.

2.A.2 Estimating $c = 1_N' \Sigma^{-1} \mu$

By assuming independent multivariate normally distributed asset returns with parameters μ and Σ and $\widehat{\mu}$ and $\widehat{\Sigma}$ being their maximum likelihood estimators, Kan and Smith (2008) show that the sample estimator $\widehat{c} = 1_N' \widehat{\Sigma}^{-1} \widehat{\mu}$ is biased. Also, $\widehat{c}$ may take negative values when theoretically unjustified, i.e., $c = 1_N' \Sigma^{-1} \mu = \mu_{\mathrm{MINV}} / \sigma^2_{\mathrm{MINV}} > 0$ should hold. One simple way to improve the sample estimator of c is to adjust its corresponding unbiased estimator $\widehat{c}_{\mathrm{u}}$ such that it never falls below a predetermined threshold value of $c_{\min}$. We therefore suggest to estimate c via

$$\widehat{c}_{\min} = \max \left\{ \widehat{c}_{\mathrm{u}}, c_{\min} \right\}, \quad \text{where} \quad \widehat{c}_{\mathrm{u}} = \frac{T - N - 2}{T} \, 1_N' \widehat{\Sigma}^{-1} \widehat{\mu}. \tag{2.14}$$

A second possibility to ensure positive estimates of c is to employ a penalized maximum likelihood approach. Kan and Smith (2008) show that $\widehat{c}_{\mathrm{u}}$ follows a non-standard distribution with

$$\mathbb{V}\mathrm{ar}\left(\widehat{c}_{\mathrm{u}}\right) = \frac{(T - N - 2) \left(\theta^2 + \frac{T-2}{T}\right) 1_N' \Sigma^{-1} 1_N + (T - N) c^2}{(T - N - 1)(T - N - 4)}. \tag{2.15}$$

Our own simulations indicate that the distribution of $\widehat{c}_{\mathrm{u}}$ can be sufficiently well approximated by a normal distribution. By using the improved estimator for θ^2 given in Kan and Zhou (2007) and the unbiased estimator $\frac{T-N-2}{T} 1_N' \widehat{\Sigma}^{-1} 1_N$ for $1_N' \Sigma^{-1} 1_N$ in (2.15), we estimate c by an approximate penalized maximum likelihood approach via

$$\widehat{c}_{\mathrm{PML}} = \arg\max_{c} \, \log f\left(\widehat{c}_{\mathrm{u}} \mid c\right) + \log \mathrm{prior}(c), \tag{2.16}$$

where $f\left(\widehat{c}_{\mathrm{u}} \mid c\right)$ is the density of a normal distribution with mean c and variance as in (2.15) and prior(c) is a prior distribution on c. The intuition here is that the penalized maximum likelihood estimator can be interpreted as a maximum a posterior probability estimator, i.e., the mode of the posterior distribution in the Bayesian sense, which is closely related to typical Bayesian estimation methods (see, e.g., Robert (2007)). When we specify a non-negative prior distribution for c it is ensured that $\widehat{c}_{\mathrm{PML}}$ is strictly positive as desired. For the prior we specify a lognormal distribution with a mode of 5 and a standard deviation of the (correspondingly normally distributed) log variable of 1. The penalized maximum likelihood estimator has the

[6] The motivation stems from the insight that $(\widehat{\mu} - \widetilde{\mu}\, 1_N)' \, (\widehat{\mu} - \widetilde{\mu}\, 1_N)$ follows a generalized χ^2-distribution. We first scale it to X, then apply the estimator proposed in Kubokawa et al. (1993) for the parameter of the non-central χ^2-distribution. Our resulting estimator often is an improvement upon the plug-in estimator in terms of bias and mean-squared error.

feature that estimates are shrunk towards the mode of the prior distribution. The amount of shrinkage is determined by the standard deviation of the prior distribution. The choice of the two prior parameters is of course to some extent arbitrary but the method has the advantage that there is no strict (positive) lower bound on the estimates. A sensibly large standard deviation of the prior distribution has the effect that the shrinkage effect is not too strong and the prior distribution is not overly dominant.

Chapter 3

Estimation of large mean-variance efficient portfolios using machine learning methods

The content of this chapter corresponds to the following working paper:
Kircher, F. (2023). Estimation of large mean-variance efficient portfolios using machine learning methods. Working paper, University of Regensburg.

Abstract
We study the implementation and the economic value of large mean-variance efficient portfolios making use of firm-specific characteristics and machine learning methods in order to deal with parameter estimation risks. We show that our estimated efficient portfolios achieve favorable out-of-sample performance in a comparative empirical evaluation. While allowing for nonlinearities improves portfolio performance by 8–10%, our best performing model, an ensemble of linear model, boosted regression trees and neural network, exhibits annualized Sharpe ratios of 2.92 and 2.32 for US and European stock market indices generating annualized alphas of 42 and 33 percent with respect to the Fama-French six-factor model. Accounting for transaction costs of 50 bps, the economic value remains significant as investors would require per annum 25 and 19 percent in addition to the respective market index return to be indifferent to our mean-variance optimized portfolio strategies.

Keywords: portfolio optimization; mean-variance; machine learning; firm characteristics; estimation risks.

JEL classification: G11, G12.

3.1 Introduction

The traditional approach of mean-variance optimal portfolio choice (Markowitz (1952)) requires the specification of expected returns and covariance matrix for all assets. This task is particularly challenging for large portfolios, e.g., a portfolio of all stocks in a broad market index, because of the huge number of parameters that needs to be estimated. Solely using historical return data to estimate first and second return moments leads to large losses in terms of out-of-sample portfolio performance (Kan and Zhou (2007)). Estimated mean-variance optimal portfolios regularly underperform equally-weighted portfolios in practical applications due to parameter estimation risks that are particularly severe for large numbers of assets, see, e.g., DeMiguel et al. (2009). Therefore, Brandt et al. (2009) propose parametric portfolio policies for individual stocks aiming to reduce the dimensionality of the optimization problem and parameter estimation risks by modeling portfolio weights as a linear function of firm-specific characteristics. Portfolio weights are effectively determined by mean-variance optimization on characteristics-managed portfolios and thereby implicitly capture the relation between the joint distribution of returns and firm-individual characteristics, see also DeMiguel et al. (2020). However, because characteristics directly determine individual portfolio weights, parametric portfolio policies do not possess the flexibility of fully specifying the risk and reward sides of the portfolio optimization problem on the stock-level.

Our approach in this paper differs because we take advantage of separately modeling individual stocks' expected returns and covariance matrix, taking into account that both return moments impose distinct requirements on model architectures. The key approach is to employ fine-tuned methods for the estimation of return moments with regard to mean-variance portfolio optimization. Specifically, we use machine learning methods like linear ridge regression, regression tree ensembles and artificial neural networks to predict expected stock returns from monthly firm characteristics. These machine learning methods are tuned to maximize the resulting portfolio's out-of-sample Sharpe ratio. For the covariance matrix we employ a shrinkage estimator based on a single-index market model estimated from daily return data. Most importantly, we demonstrate that, through the efficient combination of characteristics and covariance matrix information, our implemented large mean-variance efficient portfolios are capable of outperforming parametric portfolio policies, minimum variance portfolios and equally-weighted portfolios. An empirical out-of-sample analysis on broad US and European stock market indices reveals significant economic value, i.e., our optimized portfolios exhibit favorable annualized Sharpe ratios of up to 2.92 generating alphas of up to 42 percent per annum with respect to the Fama-French six-factor model. The estimated efficient portfolios can thus be relevant for asset pricing because the efficient portfolio directly determines the stochastic discount factor which prices all considered assets, while the sizeable outperformance of our estimated mean-variance efficient portfolios cannot be explained by the six-factor model.

One reason for the practical challenges of large portfolio optimization is that the low signal-to-noise ratio of realized stock returns makes it difficult to accurately estimate the vector of expected returns. However, anchoring expected returns to firm-individual characteristics improves their out-of-sample cross-sectional predictability as, e.g., in Lewellen (2015) and Gu et al. (2020), where machine learning methods are found useful to capture the complex nonlinear relationships. On the other hand, the covariance matrix can be successfully estimated for large asset universes based on historical return data alone. While the sample covariance matrix is inappropriate in large dimensions, shrinkage estimators perform favorably for the purpose of portfolio optimization (Ledoit and Wolf (2017)). Despite relying on monthly forecasts of expected returns as well as monthly updates of optimal portfolio weights, we use higher frequency daily data to estimate the covariance matrix via a shrinkage estimator. The estimator is based on the single-index model because, e.g., Chan et al. (1999) show that a single factor captures most of the relevant covariance structure of stock returns for portfolio optimization. We demonstrate empirically that the higher frequency return data allows to obtain more precise estimates of the covariance matrix for the mean-variance optimal portfolio choice problem.

While in the context of asset pricing nonlinear models and machine learning methods have been used to model expected returns and construct long-short decile spread portfolios from expected return estimates, e.g., Freyberger et al. (2020) and Gu et al. (2020), it remains important to distinguish between an asset's position in a long-short spread portfolio and its position in the mean-variance efficient portfolio. A stock's characteristics may predict high expected returns resulting in a long position in the spread portfolio. The stock, however, does not necessarily take up a significant position in the mean-variance efficient portfolio because a possible increase in portfolio variance can offset the effect of an increase in expected portfolio return. Because long-short spread portfolios most likely do not lead to mean-variance efficient portfolios, see, e.g., Daniel et al. (2020), we consider it important to examine the value of expected return predictions in the context of mean-variance optimal portfolio choice which is taking portfolio risk into account.

The goal of this study is to evaluate the out-of-sample performance and determine the economic value of estimated large mean-variance efficient portfolios in a comprehensive empirical analysis. Our contribution to the literature is fourfold. First, we show that there is substantial value in using firm characteristics to predict expected stock returns for large mean-variance optimal portfolio choice. Going beyond existing evidence from parametric portfolio policies, our key contribution is that our optimized portfolios achieve more than 50 percent higher out-of-sample Sharpe ratios and we are the first to show that traditionally estimated large mean-variance efficient portfolios are capable of outperforming parametric portfolio policies. In a comparative empirical analysis we obtain annualized Sharpe ratios between 2.32 and 2.92 for our best performing methods on three considered investment universes, while the out-of-sample Sharpe ratios of parametric portfolios are between 1.14 and 1.84. Against this, minimum variance and equally-weighted portfolios obtain Sharpe ratios between 0.41 and 0.86 over the 1995–2020 (US) and 2000–2020 (Europe) out-of-sample periods. Our datasets comprise investment universes based on stock market indices of large to small cap companies, i.e., the S&P 500, S&P 1500 and STOXX Europe 600, on two important stock markets, i.e., the US and Europe. Because we focus on these indices, we ensure that we solely invest in liquid stocks in our empirical analysis. Analogously to active fund managers which are often benchmarked against passive indices, we obtain a direct comparison for optimized portfolios. Our results have implications for asset pricing, in particular, the construction of the stochastic discount factor. We find that the Fama-French six-factor model is unable to explain the outperformance of our estimated mean-variance efficient portfolios which generate annualized alphas between 32 and 42 percent based on the considered stock market indices.

Second, we contribute to the literature on portfolio optimization that considers expected stock returns estimated from firm characteristics. The literature is scarce in this respect and Haugen and Baker (1996) is a notable exception. We are first to evaluate different machine learning methods in this context, i.e., for predicting expected returns for the purpose of large mean-variance optimal portfolio choice. We find that with our data a linear ridge regression model is competitive with nonlinear models based on machine learning methods such as regression tree ensembles and artificial neural networks. Our empirically best performing prediction model is an ensemble of linear and nonlinear models comprising gradient boosted regression trees and neural networks. We observe that the ensemble model's relative outperformance in long-short spread portfolios with respect to the linear model cannot be fully maintained when we turn to the mean-variance optimized portfolios. The gain in terms of out-of-sample Sharpe ratio of the ensemble compared to the linear model amounts to 8–10% based on our three considered investment universes.

Third, estimation risks in optimized portfolio weights, that arise from using estimates of expected returns and covariance matrix, lead to excessive out-of-sample variance. Estimated optimal portfolios with a constraint on portfolio variance no longer satisfy the prespecified risk limit on out-of-sample data. We extend the mean-variance optimization approach to incorporate estimation risks from optimized portfolio weights. Via the jackknife method, see, e.g., Basak et al. (2009), applied to the covariance matrix we manage to take estimation risks into account in order to construct mean-variance efficient portfolios that conform to the prespecified risk limit. Application of the jackknife method results in realized out-of-sample portfolio volatilities that are roughly consistent with the prespecified risk limit. On the two datasets based on S&P 500 and STOXX Europe 600, portfolio volatilities are within a percentage point of the specified volatility target of 15 percent per annum and within three percentage points of the target on the larger S&P 1500 dataset. We are thus successful in controlling estimation risks for optimized portfolios targeting a specific portfolio risk limit but there remains some room for improvement.

Fourth, we document high portfolio turnover for our estimated mean-variance efficient portfolios. Portfolio turnover amounts to 500 to 770 percent per month which would induce excessive transaction costs when implementing the portfolio strategies in practice. We therefore contribute by analyzing turnover-constrained strategies where portfolio weights are partially rebalanced each month if the required turnover exceeds a given limit. We find that constraining the strategies to a monthly portfolio turnover of 100 percent results in only small losses in out-of-sample Sharpe ratios. We obtain gains in net Sharpe ratios of up to 280 percent compared to the equally-weighted portfolios even after accounting for proportional round-trip transaction costs of 0.5%. These results emphasize the economic value of the portfolio strategies. We find that investors would require per annum 25 percent on the S&P 1500 and 19 percent on the STOXX Europe 600 in addition to the respective market index return in order to be indifferent to our mean-variance optimized portfolios. An additional feature of the turnover-reduced portfolios is that their out-of-sample volatilities are considerably lower, i.e., mostly below 13 percent per annum with a targeted value of 15 percent for our optimized portfolios.

The following Section 3.2 describes the relation of this paper to the literature. Section 3.3 presents the datasets on which our empirical analysis is built. Section 3.4 introduces the methodology and models for estimating expected returns and covariance matrix and describes the portfolio optimization strategy. Section 3.5 discusses the results of an empirical application to three index-based investment universes. Section 3.6 concludes the paper.

3.2 Related literature

The literature has addressed the problem of mean-variance portfolio optimization for large numbers of individual stocks by parametric portfolios or, as in Ao et al. (2019), by considering the Fama-French factors and enforcing sparsity on the weights of individual stocks via the lasso. Parametric portfolios incorporate firm characteristics by modeling portfolio weights as a linear function of firm-specific variables. The linear specification implies that portfolio optimization is performed on characteristics-managed portfolios whose weights are determined by the firms' exposures to the respective characteristic. Framing the optimization problem in terms of characteristics-managed portfolios, where the number of characteristics is smaller than the number of assets, reduces the dimensionality of the problem and mitigates estimation risks for optimal portfolio weights. Initiated by Brandt et al. (2009), the method of parametric portfolio policies has been refined and applied in various contexts. While Brandt et al. (2009) make use of the size, book-to-market and momentum characteristics, Hand and Green (2011) apply the method to accounting-based characteristics, i.e., accruals, change in earnings and asset growth. Hjalmarsson and Manchev (2012) show that for a mean-variance utility function parametric portfolio policies are equivalent to mean-variance optimization on characteristics-managed portfolios. DeMiguel et al. (2013) find that incorporating stock option-implied information improves the out-of-sample performance of parametric portfolios. Ammann et al. (2016) introduce a leverage constraint to parametric portfolio optimization helping to reduce estimation risks and transaction costs. DeMiguel et al. (2020) show that the number of jointly significant characteristics for an investor's optimal portfolio increases when transaction costs are incorporated into the portfolio optimization problem.

Furthermore, the paper is related to literature on out-of-sample tests of cross-sectional prediction models for expected returns. Haugen and Baker (1996) and Hanna and Ready (2005) use estimates of individual stock's expected returns from Fama-MacBeth regressions to evaluate the profitability of long-short portfolios. Adding to that, Lewellen (2015) studies the cross-sectional properties and predictive accuracy of Fama-MacBeth regressions using a maximum of 15 firm characteristics. Gu et al. (2020) employ machine learning methods to predict expected returns from a set of 94 firm characteristics. They compare cross-sectional predictive accuracy among a number of methods such as penalized linear regression, regression tree ensembles and artificial neural networks. The nonlinear models display a favorable out-of-sample performance in terms of profitable long-short portfolios. But their usefulness for mean-variance optimal portfolio choice is not evaluated. Because portfolio risk, i.e., the (co-)variances of decile portfolios, is not taken into account for portfolio formation, the construction of long-short portfolios from cross-sectional decile splits of expected returns does not lead to mean-variance efficient portfolios (Daniel et al. (2020)).

Most closely related to our approach are Haugen and Baker (1996) conducting an out-of-sample analysis of mean-variance optimized portfolios where expected returns are estimated from firm characteristics via a linear regression model. They find that optimized portfolios which incorporate weight and turnover constraints dominate the market portfolio. Chan et al. (1999) compare the out-of-sample performance of covariance matrix prediction models for minimum variance portfolios. Their analysis reveals that a one-factor model is sufficient for the estimation of the covariance matrix. Ledoit and Wolf (2003, 2017) and De Nard et al. (2021) achieve favorable out-of-sample performance with shrinkage covariance matrices in conjunction with a one-factor market model.

Adding to the above literature, we find that for mean-variance portfolio optimization the full specification of expected returns and covariance matrix for all assets entails significant advantages over the more restrictive specification of parametric portfolio weights. The increased flexibility allows to employ fine-tuned models for the respective return moments. Firm characteristics are useful predictors for expected returns while the covariance matrix may be estimated from daily returns with a shrinkage estimator based on the single-index model. Employing flexible models such as neural networks for expected return prediction improves the out-of-sample performance of our mean-variance optimized portfolios. These results can be of use in an asset pricing context because of the direct relation of the mean-variance efficient portfolio to the projected stochastic discount factor on the considered asset returns (see Cochrane (2005) and DeMiguel et al. (2020)). Various papers employ machine learning methods such as regularization techniques (Kozak et al. (2020), Freyberger et al. (2020), Feng et al. (2020)), principal component analysis (Kelly et al. (2019), Lettau and Pelger (2020), Kim et al. (2021)), regression trees (Moritz and Zimmermann (2016), Bryzgalova et al. (2023)) and artificial neural networks (Gu et al. (2021), Chen et al. (2024)) for asset pricing.

3.3 Data

The study concentrates on stocks that are constituents of three major stock market indices: the S&P 500, S&P Composite 1500 and STOXX Europe 600. The two S&P indices represent investment universes of large US companies and large to small cap US companies, respectively, whereas the STOXX index covers large to small cap companies of the European stock market. The choice to consider only stocks for portfolio optimization that are index constituents implies that we preclude those for investment that have low stock prices and are less liquid. End of month constituent lists are available from Refinitiv Datastream from 09/1989 for the S&P 500, from 12/1994 for the S&P 1500 and from 08/1999 for the STOXX Europe 600, all until 12/2020. Firm-individual data is retrieved from the Refinitiv Datastream and Refinitiv Worldscope databases. Earliest firm-specific annual accounting data are available starting in 1980. Following Lewellen (2015) we assume that the annual accounting data are known four months after the firm's fiscal year end. For each firm and each month in the data sample we calculate 27 firm-specific variables which are detailed in Table 3.A.1 in the appendix, e.g., some of the characteristics are logarithmized and therefore take negative values, e.g., the volatilities. The set of variables is a selection of characteristics considered in previous literature on cross-sectional return prediction and asset pricing, e.g., in Lewellen (2015) and Freyberger et al. (2020). The set comprises past return characteristics, such as momentum and reversal, stock- and market-risk-based characteristics,

Table 3.1: Descriptive statistics of 27 firm variables for constituents of the S&P 500, S&P 1500 and STOXX Europe 600 indices.

	S&P 500			S&P 1500			STOXX Europe 600		
Variable	N	Mean	Std	N	Mean	Std	N	Mean	Std
Size	498.79	8.82	1.06	1405.18	7.28	1.46	566.54	8.31	1.11
B/M	467.88	−1.05	0.76	1296.41	−0.92	0.75	527.50	−0.87	0.86
$\text{Ret}_{1,0}$	498.77	0.01	0.07	1403.33	0.01	0.09	565.76	0.01	0.07
$\text{Ret}_{12,2}$	495.76	0.13	0.26	1382.65	0.14	0.32	554.36	0.13	0.27
$\text{Ret}_{12,7}$	495.78	0.07	0.18	1382.70	0.08	0.22	554.38	0.07	0.19
$\text{Ret}_{36,13}$	487.56	0.35	0.48	1327.69	0.38	0.59	521.27	0.37	0.52
$\text{Ret}_{60,13}$	477.14	0.85	0.96	1258.09	0.90	1.14	484.80	0.87	1.00
DY	498.90	0.02	0.02	1404.85	0.02	0.02	566.43	0.03	0.02
Vol	477.08	−2.49	0.31	1257.92	−2.35	0.36	484.75	−2.46	0.29
Beta	477.08	1.04	0.48	1257.92	1.08	0.52	488.16	1.05	0.44
IVol	477.08	−2.67	0.31	1257.92	−2.50	0.37	488.16	−2.66	0.30
$\text{Vol}_{\text{daily}}$	498.58	−2.59	0.37	1401.27	−2.44	0.44	562.51	−2.60	0.37
$\text{Beta}_{\text{daily}}$	498.58	1.01	0.61	1401.27	0.99	0.73	562.51	0.89	0.60
$\text{IVol}_{\text{daily}}$	498.58	−2.80	0.41	1401.27	−2.62	0.48	562.51	−2.80	0.40
Turnover	498.07	−1.81	0.82	1391.61	−1.88	0.94	550.23	−2.45	1.07
Investment	479.29	0.10	0.17	1321.65	0.13	0.21	553.27	0.12	0.21
Debt/Price	459.04	−1.49	1.41	1199.43	−1.60	1.61	524.58	−1.20	1.71
Sales/Price	476.21	−0.34	0.88	1314.94	−0.27	0.97	533.58	−0.25	1.07
SalesGrowth	479.46	0.07	0.13	1325.31	0.09	0.16	551.80	0.08	0.16
Earnings/Price	475.58	0.05	0.05	1312.94	0.04	0.06	534.65	0.06	0.07
EarningsGrowth	478.71	0.25	1.15	1320.78	0.25	1.30	553.20	0.27	1.15
GrossProfit	407.20	0.39	0.21	1141.18	0.41	0.23	456.39	0.29	0.18
CF/Price	476.10	0.12	0.09	1313.34	0.12	0.09	533.45	0.14	0.14
CashRatio	462.03	−3.15	1.29	1215.00	−3.11	1.37	546.17	−2.81	1.11
Accruals	403.12	0.01	0.17	1110.64	0.03	0.18	420.31	0.01	0.21
Issues	487.56	0.03	0.17	1327.57	0.07	0.18	521.11	0.10	0.21
MAX	498.84	0.04	0.02	1403.70	0.05	0.03	565.75	0.04	0.02

Note: The table reports time-series averages of three cross-sectional statistics: the number of firms N with available characteristics data, the cross-sectional mean and standard deviation of characteristics. Characteristics are cross-sectionally winsorized on the 3% level each month. The sample periods are from 12/1982 for the S&P indices and from 12/1989 for the STOXX index ending in 12/2020. Each month we consider firms that are constituents of the respective index; for months in which constituent lists are not available we take the constituents of the first available list. For US stocks, market beta is calculated with data on the market index and risk-free return from Kenneth French's data library. For European stocks, we take the perspective of a European investor for whom all stock returns are denominated in euro. The market index is the STOXX Europe 600 which is available from 12/1986. We therefore calculate beta and idiosyncratic volatility based on 36 monthly returns for the first two years in the sample. The European risk-free rate is based on the 3-month FIBOR until 1998 and on the 1-month euro OIS rate afterwards.

e.g., beta, size, book-to-market and volatility, and investment- as well as profitability-based characteristics. The characteristics data are winsorized on the 3% level following DeMiguel et al. (2013), i.e., observations below the 3%-quantile or above the 97%-quantile are set equal to the respective quantile.

Table 3.1 reports descriptive statistics of the firm-specific variables, i.e., time-series averages of the monthly number of firms who are constituents of the respective indices and have available characteristics data, as well as monthly cross-sectional means and standard deviations of characteristics. The data samples are from 12/1982 for the two S&P indices and from 12/1989 for the STOXX index. Because our data samples start before the first constituent lists are available, we use the first list's constituents in this case. The monthly number of firms therefore increases by construction until the first list of an index is available. For most characteristics the average monthly number of firms from the S&P 500 is between 470 and 500, for the accruals and gross profitability characteristics, which display the lowest availability, it is around 400. This pattern repeats for the S&P 1500 and STOXX 600 indices where we have around 1300 and 500 firms for most characteristics. Abstracting from different sample periods and firm selection, the descriptive statistics of characteristics in Table 3.1 are comparable in size with the statistics reported in Lewellen (2015) and Freyberger et al. (2020). The cross-sectional means and standard deviations also appear qualitatively similar across the three stock market indices. The cross-sectional variation of characteristics increases with the larger number of firms in the S&P 1500 dataset as one would expect.

Figure 3.1 plots the number of firms which have fully available characteristics data over time. The monthly number of firms is therefore reduced with regard to the metrics reported in Table 3.1 which provides information on a variable by variable basis. Figure 3.1 represents the monthly number of firms for which we are able to predict expected returns and include in the monthly constructed portfolios. The number of firms strongly increases until the first constituent lists are available, especially for the S&P 1500 and STOXX Europe 600 indices. The number of firms in the constructed portfolios then is around 350 for the S&P 500 and around 1000 for the S&P 1500 datasets. For the STOXX Europe 600 dataset the number of firms in the portfolios increases from around 300 in 01/2000 to over 400 in the 2010s.

3.4 Methodology

In this section we describe specifications of expected return prediction models based on individual firm characteristics. We present the approach for mean-variance portfolio optimization taking parameter estimation risks into account and describe the procedures to estimate expected returns

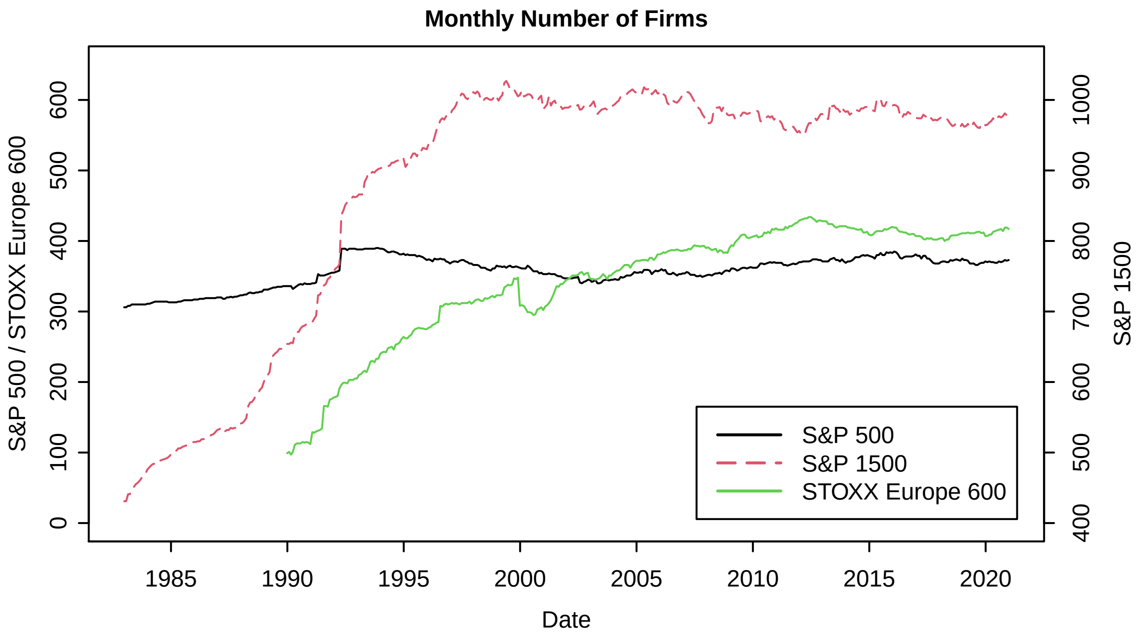

Figure 3.1: Monthly number of index constituents having fully available characteristics data in the respective data samples. Left y-axis belongs to the S&P 500 and STOXX Europe 600, right y-axis to the S&P 1500 data sample. Constituent lists are available from 09/1989 for S&P 500, from 12/1994 for S&P 1500 and from 08/1999 for STOXX Europe 600. The data samples start in 12/1982 for the S&P indices and in 12/1989 for the STOXX index.

via machine learning methods as well as the covariance matrix via a shrinkage estimator.

3.4.1 Prediction models for expected excess returns

Let $R_{i,t}$ denote the excess stock return (over the risk-free asset) of firm i over month t, i.e., from end of month $t-1$ to end of month t, and $R_{m,t}$ the average excess return of firms in a stock market index, i.e., the S&P 500, S&P 1500 or the STOXX Europe 600. We model the (conditional) expected return of firm i, $\mu_{i,t}$, in excess of the expected average return of firms in the index, $\mu_{m,t}$, as a function f of firm characteristics:

$$\mathbb{E}_{t-1}\left(R_{i,t}\right) - \mathbb{E}_{t-1}\left(R_{m,t}\right) = \mu_{i,t} - \mu_{m,t} = f(X_{i,t-1}), \tag{3.1}$$

where $\mathbb{E}_{t-1}$ is the expected value conditional on all information available until the end of month $t-1$, e.g., the characteristics for month $t-1$. $X_{i,t-1}$ denotes the vector of cross-sectionally winsorized and standardized firm characteristics for firm i at the end of month $t-1$. The model in (3.1) establishes a predictive relationship between firm characteristics and (conditional) expected excess returns. It implies that firm i's characteristics predict its expected return in excess of the average market return over the next month. Firm-specific data from months before $t-1$ as well as the dynamics of firm characteristics are not represented in (3.1). This restriction is not detrimental because older characteristics should matter less and dynamics may be incorporated into variables in $X_{i,t-1}$. Because (3.1) constitutes a cross-sectional model, the function f describing the relation of expected returns and firm characteristics holds for each firm i. To account for some variation over time in f we periodically re-fit the model to the data using a rolling window which is a common approach in the literature, see, e.g., Gu et al. (2020).

We consider different specifications of the function f. The benchmark specification of f is a linear model implying that

$$\mu_{i,t} - \mu_{m,t} = \theta_0 + \theta' X_{i,t-1}, \tag{3.2}$$

where θ_0 denotes the intercept and θ is the parameter vector whose dimension is equal to the number of firm characteristics. The linear benchmark specification is, in our case, not designed to capture nonlinear or interaction effects of firm characteristics on expected returns, which could be achieved by including quadratic or cubic terms of $X_{i,t-1}$ in (3.2).

The nonlinear specifications of f are based on regression tree ensembles and artificial neural networks. We distinguish two ways of constructing a regression tree ensemble. One way is to average over a number of individual regression trees built in parallel on subsamples or resamples of the data which is called bootstrap aggregating or bagging. The other is to iteratively fit regression trees to the residuals, i.e., errors, of the previously fitted trees, called gradient boosting in which each iteration may also be applied to subsets of the data sample. The expected return model based on either of the regression tree ensembles can be written as

$$\mu_{i,t} - \mu_{m,t} = \sum_{j=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{K}_j} \theta_{j,k} \, \mathbb{1}\{X_{i,t-1} \in C_{j,k}\}, \tag{3.3}$$

where $\mathcal{N}$ is the number of trees, $\mathcal{K}_j$ the number of terminal nodes on the j-th tree, $C_{j,k}$ a partition of the input space and $\theta_{j,k}$ the score connected with the k-th terminal node on the j-th tree. The indicator function in (3.3) partitions the input space of firm characteristics. If the partition $C_{j,k}$ contains the values of firm variables $X_{i,t-1}$, then the parameter value $\theta_{j,k}$ is added as one of the summands that form the prediction. The model thus represents a combination of piecewise constant functions and thereby is able to capture nonlinear and interaction effects.

The feedforward neural network model consists of a sequence of layers which are consecutively combined such that each layer's inputs are given by the outputs of the previous layer. The first layer processing the input variables is called input layer, the following layers are called hidden layers and the final layer assembling the outputs of the last hidden layer is called output layer. For each hidden layer the number of units or nodes is specified to determine the complexity of the layer and, together with the number of hidden layers, the complexity of the network. Each unit of a layer is connected to another unit of the following layer through an individual weight. In our empirical application we found a neural network with one hidden layer to be sufficient which is why we define the neural network model with only one hidden layer. The model can be written as

$$\mu_{i,t} - \mu_{m,t} = \theta_0^{(1)} + \sum_{j=1}^{\mathcal{N}} \theta_j^{(1)}\, h\left(\theta_{0,j}^{(0)} + \theta_j^{(0)\,\prime} X_{i,t-1}\right), \tag{3.4}$$

where $\theta_0^{(1)}$ is the intercept of the output layer, $\mathcal{N}$ denotes the number of units of the hidden layer and $\theta_j^{(1)}$ the weight of the output layer corresponding to unit j of the hidden layer. h denotes the activation function which transforms the weighted sum of inputs of a unit in the hidden layer. $\theta_{0,j}^{(0)}$ is the intercept and $\theta_j^{(0)}$ the vector of weights corresponding to unit j of the hidden layer. The activation function h is, in our case, the so-called rectified linear unit (ReLU), defined as $h(x) = \max\{0, x\}$. The neural network model thus constitutes a combination of $\mathcal{N}$ piecewise linear functions and thereby is able to represent nonlinear and interaction effects.

3.4.2 Portfolio optimization with a risk limit and parameter estimation risks

The portfolio optimization task that we consider is based on Markowitz (1952) and aims at maximizing the expected portfolio return for a given risk constraint σ^2 on the portfolio variance. The optimization problem can be written as

$$\max_w w'\mu_t \quad \text{s.t.} \quad w'\Sigma_t w \leq \sigma^2, \quad \text{where} \quad w_t = \frac{\sigma}{\sqrt{\mu_t'\Sigma_t^{-1}\mu_t}}\,\Sigma_t^{-1}\mu_t \tag{3.5}$$

is the solution and vector of optimal portfolio weights. $\mu_t = (\mu_{1,t}, \ldots, \mu_{N_t,t})'$ denotes the vector of expected returns of N_t firms for month t and Σ_t the covariance matrix of N_t stock returns. Having defined models for expected returns $\mu_{i,t}$ based on firm characteristics, we have to estimate the parameters θ of the models in (3.2)–(3.4) based on a finite sample of data in order to determine optimal portfolio weights w_t in practice. This procedure inevitably introduces estimation errors so that estimated portfolio weights $\widehat{w}_t$ do not coincide with w_t and thus are no longer optimal. In particular, the prespecified risk constraint σ^2 will not be satisfied with parameter estimation errors in $\widehat{w}_t$. We extend the problem in (3.5) to estimated portfolio weights and define the optimization problem taking parameter estimation risks into account as

$$\max_{c} \mathbb{E}\left[\widehat{w}_t(c)'\mu_t\right] \quad \text{s.t.} \quad \mathbb{E}\left[\widehat{w}_t(c)'\Sigma_t \widehat{w}_t(c)\right] \leq \sigma^2, \tag{3.6}$$

where $\widehat{w}_t(c) = c\,\widehat{\Sigma}_t^{-1}\widehat{\mu}_t$ parameterizes the estimated portfolio weights similar to (3.5). The objective formulated in (3.6) is to maximize the expected out-of-sample portfolio return for a given risk constraint on the expected out-of-sample portfolio variance. The solution to (3.6) is

$$c_t^* = \sigma\left(\mathbb{E}\left[\widehat{\mu}_t'\widehat{\Sigma}_t^{-1}\Sigma_t\widehat{\Sigma}_t^{-1}\widehat{\mu}_t\right]\right)^{-1/2} \tag{3.7}$$

which reduces to $\frac{\sigma}{\sqrt{\mu_t'\Sigma_t^{-1}\mu_t}}$ if there is no parameter uncertainty as in (3.5).

In practice we have to estimate the expected value in c_t^* in (3.7). This could be done via resampling methods, but the excessive runtime from re-estimating the expected return models from Section 3.4.1 via machine learning models renders the task barely feasible. In order to circumvent the issue, we assume that the estimators $\widehat{\mu}_t$ and $\widehat{\Sigma}_t$ are independent. An assumption that would be satisfied if asset returns were iid multivariate normally distributed and $\widehat{\mu}_t$ and $\widehat{\Sigma}_t$ are typical sample estimators. It then holds that $\mathbb{E}\left[\widehat{\mu}_t'\widehat{\Sigma}_t^{-1}\Sigma_t\widehat{\Sigma}_t^{-1}\widehat{\mu}_t\right] = \mathbb{E}\left[\widehat{\mu}_t'\,\mathbb{E}\left[\widehat{\Sigma}_t^{-1}\Sigma_t\widehat{\Sigma}_t^{-1}\right]\widehat{\mu}_t\right]$ and $\mathbb{E}\left[\widehat{\Sigma}_t^{-1}\Sigma_t\widehat{\Sigma}_t^{-1}\right]$ can be estimated via the jackknife method, see, e.g., Basak et al. (2009). Denoting the resulting jackknife estimator by $\widehat{\Sigma}_{t,\text{JK}}^{-1}$, we estimate c^* in (3.7) via $\widehat{c_t^*} = \sigma\left(\widehat{\mu}_t'\widehat{\Sigma}_{t,\text{JK}}^{-1}\widehat{\mu}_t\right)^{-1/2}$ to obtain estimated portfolio weights

$$\widehat{w}_t(\widehat{c_t^*}) = \frac{\sigma}{\sqrt{\widehat{\mu}_t'\widehat{\Sigma}_{t,\text{JK}}^{-1}\widehat{\mu}_t}}\,\widehat{\Sigma}_t^{-1}\widehat{\mu}_t. \tag{3.8}$$

In (3.8), $\widehat{\mu}_t = \widehat{\mu}_{m,t}\,1_{N_t} + (\widehat{f}(X_{1,t-1}), \ldots, \widehat{f}(X_{N_t,t-1}))'$ are the predicted expected returns for the N_t firms in the portfolio for month t. $\widehat{\mu}_t$ consists of two components. The predicted market return $\widehat{\mu}_{m,t}$ is the same for each asset in the portfolio with 1_{N_t} denoting a vector of ones of length N_t. The magnitude of $\widehat{\mu}_{m,t}$, being represented in the numerator as well as in the denominator in (3.8), has only limited influence on the magnitude and spread of portfolio weights $\widehat{w}_t$ across assets. $\widehat{\mu}_{m,t}$, which is the grand mean of the assets' mean returns, affects all assets equally as an add-on to individual expected excess returns, thus not altering their relative ranking as argued in Jorion (1985). We therefore do not specify a complicated model for the expected market return $\mu_{m,t}$, but estimate $\mu_{m,t}$ as the grand sample mean of past constituent stock returns using all data from the beginning of the respective sample until month $t-1$.

3.4.3 Estimating the models via machine learning methods

We aim to compare the different models and methods from Section 3.4.1 for the prediction of firm-individual expected stock returns in (3.1). This requires estimating the function f which is determined by the parameters of the models given in (3.2)–(3.4). In order to prevent overfitting the models to finite data samples, we adopt regularization techniques. The parameter vector (θ_0, θ') of the linear model in (3.2) is estimated by minimizing the sum of squared errors plus a penalty term based on the squared ℓ_2-norm of θ, which is also referred to as a ridge regression:

$$\min_{\theta} \sum_{s=t-T}^{t-1} \sum_{i=1}^{N_s} \left(R_{i,s} - R_{m,s} - \theta_0 - \theta' X_{i,s-1}\right)^2 + \lambda\, \theta'\theta. \tag{3.9}$$

In (3.9), parameter estimation is performed at the beginning of month t where the previous T months of data are used for estimation. The parameter λ determines the strength of regularization. It penalizes the sum of squared parameter values leading to smaller and less extreme parameter estimates thus reducing the risk of overfitting the sample data. An adequate choice of λ leads to an increase in out-of-sample performance of the model. Because the optimal choice of λ is specific to each dataset, the regularization parameter is commonly selected using data-driven approaches. We describe our general approach for determining the models' hyperparameters in Section 3.5.

To estimate regression tree ensembles we employ two methods: random forests and stochastic gradient boosting. Random forests sample from the data $\mathcal{N}$ times, build a regression tree on each subsample and average the predictions of the $\mathcal{N}$ trees to arrive at a prediction of the overall model (Breiman (2001)). Regression trees are built recursively with a greedy algorithm. At each node for each considered input variable the optimal split point is determined in terms of the objective criterion, i.e., the sum of squared errors. The best splitting variable at the respective node is chosen and the parameter values assigned to the two resulting nodes or partitions of the input space are the averages of the response variable corresponding to these partitions (Hastie et al. (2009)). The algorithm proceeds with the partitioning of nodes until a specified number of depth of the tree or a minimum terminal node size (in terms of the number of observations in the respective partition) is reached.

Stochastic gradient boosting combines regression trees iteratively. It starts with sampling from the data and fits an often shallow regression tree to the first subsample. After calculating the prediction errors from the first iteration, the next boosting iteration fits a regression tree to a subsample of the previous iteration's residuals (Friedman (2001, 2002)). The algorithm proceeds until $\mathcal{N}$ boosting iterations are completed. Regularization is included by weighting each added tree with a factor $\nu < 1$ reducing the contribution of individual trees. By using the XGBoost library from Chen and Guestrin (2016) we make use of additional regularization techniques such as adding an ℓ_2-penalty term to the squared error loss with a hyperparameter λ as in (3.9) when fitting individual regression trees.

We fit the neural network model to the sample data by minimizing the sum of squared errors plus a regularization term on the network's weights, $\theta_j^{(1)}$ and $\theta_j^{(0)}$ in (3.4), similar to (3.9). The most commonly used optimization algorithm for neural networks is stochastic gradient descent where model parameters are updated iteratively based on the gradient computed with the backpropagation algorithm on subsamples of the data (Goodfellow et al. (2016)). The learning rate controls the size of parameter updates. In the Adam algorithm of Kingma and Ba (2015), which we use for minimization of the objective function, the learning rate is adapted individually to each model parameter. When the model parameters have been updated based on all subsamples of the data, the algorithm has completed an epoch. We employ dropout regularization during the fitting process emulating the training of an ensemble of models, that is, before each iteration each unit in the hidden layer is dropped with a certain probability. Hyperparameters of the neural network model that need to be determined are thus the number of units in the hidden layer, the dropout rate, the size of the subsamples, i.e., the batch size, the specification of the learning rate, the number of training epochs and the ℓ_2-regularization parameter λ.[1] Hyperparameter specification is described in more detail in Section 3.5 and in Table 3.A.2 in the appendix.

Having estimated the expected return models, an estimator $\widehat{\Sigma}_t$ for the unknown covariance matrix Σ_t of asset returns is needed to calculate portfolio weights based on (3.8). One could specify a dynamic conditional covariance model as in Engle et al. (2019), but because we primarily focus on the expected return models and their value for portfolio optimization, for simplicity, we estimate a static covariance matrix within each short estimation window of the data. The large number of assets in the portfolio relative to the number of return observations for each asset implies a huge number of parameters to be estimated for the covariance matrix. This in turn implies great estimation risks for portfolio optimization. For this reason we, firstly, use daily return data to estimate the covariance matrix and, secondly, employ a shrinkage estimator on the basis of a one-factor model because a reasonably high correlation of individual stock returns can be attributed to exposure to a common market factor. In our case, the grand mean return of firms in the portfolio serves as the single market factor. The covariance matrix

[1] We have also considered using deeper networks, i.e., more than one hidden layer in the model, but did not find any advantages in terms of out-of-sample performance on our datasets.

estimator is given as

$$\widehat{\Sigma}_t = \widehat{\Sigma}_{\mathrm{F},t} + \widehat{\Sigma}_{\mathrm{LW},t} \quad \text{with} \quad \widehat{\Sigma}_{\mathrm{F},t} = \widehat{\beta}_t \widehat{\beta}_t' \, \widehat{\sigma}^2_{m,t} \qquad (3.10)$$
$$\text{and} \quad \widehat{\Sigma}_{\mathrm{LW},t} = \widehat{\alpha}_t \, \widehat{\sigma}^2_{\varepsilon,t} \, I_{N_t} + (1 - \widehat{\alpha}_t) \, \widehat{\Sigma}_{\varepsilon,t}.$$

$\widehat{\Sigma}_{\mathrm{F},t}$ is the covariance matrix estimator based on the one-factor model. It is determined by the slope coefficients $\widehat{\beta}_t = (\widehat{\beta}_{1,t}, \ldots, \widehat{\beta}_{N_t,t})'$ of regressions of individual stock returns on the cross-sectional average return of stocks in the portfolio whose estimated variance is denoted by $\widehat{\sigma}^2_{m,t}$. $\widehat{\Sigma}_{\mathrm{LW},t}$ is the shrinkage covariance matrix estimator of Ledoit and Wolf (2004b) applied to the regression residuals whose sample covariance matrix is denoted by $\widehat{\Sigma}_{\varepsilon,t}$. $\widehat{\sigma}^2_{\varepsilon,t}$ and I_{N_t} denote the average residual variance and an $N_t \times N_t$ identity matrix yielding the shrinkage target. We use the estimator as given in Ledoit and Wolf (2004b) for the shrinkage intensity $\widehat{\alpha}_t$. Because we estimate the covariance matrix from daily data, we scale the resulting $\widehat{\Sigma}_t$ by 21 to conform to monthly data.

3.5 Empirical analysis

We conduct a comparative empirical out-of-sample analysis of mean-variance-optimized portfolios employing machine learning methods for predicting expected returns based on individual firm characteristics. This section describes the empirical approach and results of the application to three real-world equity market datasets.

3.5.1 Empirical approach

Firm characteristics are winsorized and standardized based on cross-sectional means and standard deviations on a variable by variable basis. To facilitate estimation, especially for the neural network model, we multiply the target variable by 100, i.e., we estimate the models from Section 3.4.1 on percentage returns. We divide the datasets in an initial estimation period and an out-of-sample testing period. For the S&P 500 and S&P 1500 indices the out-of-sample periods are from 01/1995 to 12/2020, for the STOXX Europe 600 index it is from 01/2000 to 12/2020. For each month within the out-of-sample period we construct a portfolio with weights as given in (3.8). Firms that are included in the portfolio for month t have to be index constituents at the end of month $t-1$ to prevent a look-ahead bias. The firms need to have fully available data on all characteristics in month $t-1$ and at least 70% of nonzero daily return data over the previous 24 months. The resulting number of firms in the portfolio is denoted by N_t. We estimate the prediction models of expected returns as described in Section 3.4.3

based on data of the $T = 144$ months, i.e., 12 years, prior to month t. For each month within the estimation window we exclusively consider index constituent firms with complete characteristics data.[2] Because monthly re-estimation is time-consuming, we re-estimate the expected return prediction models at the end of each year, when approximately 8.3% of the data for estimation are updated.

Estimating the models from Section 3.4 requires the specification of several hyperparameters. We determine the hyperparameters via an out-of-sample validation approach. The 12 years of data for estimation are divided into disjoint 12 folds where each fold contains the complete cross-sectional firm data of 12 successive months. We then estimate each model for each hyperparameter specification on 11 of the 12 folds and evaluate its performance on the omitted fold. Because the aim is to construct mean-variance optimal portfolios, we evaluate out-of-sample model performance via the Sharpe ratio. We find that optimal hyperparameters as derived from the out-of-sample validation differ when they are determined via the Sharpe ratio or the mean squared error criterion. The mean squared error criterion favors models that are strongly regularized, i.e., are more conservative than models from the Sharpe ratio criterion. We thus observe that it is important to use a criterion in the out-of-sample validation that is consistent with the objective which in our case is the maximization of the portfolio's Sharpe ratio as implied by (3.5).

For the out-of-sample validation we estimate covariance matrices $\widehat{\Sigma}_{\text{cv},t}$ for each month t of the T months in the estimation period based on the previous 24 months of daily return data using the shrinkage estimator given in (3.10). For each month in the validation fold we predict expected excess returns for each firm. The expected market return, whose out-of-sample validation is of minor importance for portfolio performance, is the same for all folds as for portfolio choice and not re-estimated. For each month t in the validation fold we calculate portfolio weights as $\widehat{w}_{\text{cv},t} = \widehat{\Sigma}_{\text{cv},t}^{-1}\,\widehat{\mu}_{\text{cv},t}$, where $\widehat{\mu}_{\text{cv},t}$ is specified as in (3.8). The portfolio weights $\widehat{w}_{\text{cv},t}$ are proportional to the optimal weights from (3.6) and (3.7) and thus achieve identical out-of-sample Sharpe ratios. For each month in the validation fold we record a portfolio return from the realized firm returns yielding in total $T = 144$ out-of-sample portfolio returns from which we calculate the Sharpe ratio in the out-of-sample validation procedure.

The optimal hyperparameters found in the out-of-sample validation procedure are used for estimation of the expected return prediction models. Because running the validation procedure

[2] For the months prior to 09/1989 for the S&P 500, 12/1994 for the S&P 1500 or 08/1999 for the STOXX Europe 600, where we do not have data on the constituents, we resort to the lists from 09/1989, 12/1994 and 08/1999, respectively. This only concerns estimation periods of the models and does not affect the out-of-sample testing periods beginning in 01/1995 for the S&P indices and 01/2000 for the STOXX index. The initial estimation period for the STOXX Europe 600 dataset is from 12/1989 until 12/1999, i.e., 10 years. It expands the following 24 months until it reaches the specified estimation window size of $T = 144$ months.

is time-consuming, we update the hyperparameters every year for the ridge regression and every five years for the gradient boosted regression trees and the neural network. The random forest which is less sensitive to hyperparameter choice receives a fixed set of hyperparameters for all datasets and time periods. Table 3.A.2 in the appendix summarizes the hyperparameter values for the four considered methods over the out-of-sample testing periods of the three datasets. In our empirical application we make predictions of expected returns for each firm and each month t in the out-of-sample testing period, yielding $\widehat{\mu}_t$, and determine the portfolio weights $\widehat{w}_t$ from (3.8). The covariance matrix $\widehat{\Sigma}_t$ is either estimated from the previous 12 years of monthly data or the previous 24 months of daily data using the shrinkage estimator based on a one-factor market model as given in (3.10). The jackknife procedure to determine $\widehat{\Sigma}_{t,\text{JK}}$ from Section 3.4.2 is conducted every month in the out-of-sample testing period. It repeatedly subsamples the return data R_{JK} of length T_{JK} leaving out one observation at a time, estimates a shrinkage covariance matrix $\widehat{\Sigma}^{-1}_{t,-j}$ on each subsample and uses the left out observation $R_{j,\text{JK}}$ to approximate $\mathbb{E}\left[\widehat{\Sigma}^{-1}_t \Sigma_t \widehat{\Sigma}^{-1}_t\right]$ via the jackknife estimator $\widehat{\Sigma}_{t,\text{JK}} = \frac{1}{T_{\text{JK}}} \sum_{j=1}^{T_{\text{JK}}} \widehat{\Sigma}^{-1}_{t,-j} (R_{j,\text{JK}} - \bar{R}_{\text{JK}})(R_{j,\text{JK}} - \bar{R}_{\text{JK}})' \widehat{\Sigma}^{-1}_{t,-j}$, where $\bar{R}_{\text{JK}}$ denotes the vector of sample means of return observations R_{JK}.[3]

We compare the portfolios based on predictions from machine learning methods with a sample mean-variance portfolio calculated from (3.8) that only uses return data, i.e., where $\widehat{\mu}_t$ in (3.8) is estimated via sample means of past $T = 144$ returns. We also construct an ensemble model that averages the expected return predictions of the linear ridge regression, gradient boosted regression trees and neural network model. As to portfolios that do not consider any estimates of expected returns we consider an equally-weighted portfolio and an estimated minimum variance portfolio which are given as

$$w_{\text{EW},t} = \frac{1}{N_t} \quad \text{and} \quad \widehat{w}_{\text{MINV},t} = \frac{\widehat{\Sigma}^{-1}_t 1_{N_t}}{1'_{N_t} \widehat{\Sigma}^{-1}_t 1_{N_t}}.$$

An important comparison is with parametric portfolio policies that use firm characteristics to directly model portfolio weights. Following Brandt et al. (2009), portfolio weights are parameterized as $w_t(\theta) = w_{b,t} + X_{t-1}\theta/N_t$, where $w_{b,t}$ are benchmark portfolio weights, i.e., weights of a value-weighted portfolio, X_{t-1} denotes a $N_t \times K$ matrix of K cross-sectionally standardized characteristics on N_t assets in the portfolio for month t and θ is a parameter vector of length K. The portfolio return $w_t(\theta)' R_t$ can be decomposed into the benchmark portfolio return

[3] As the jackknife procedure is time-consuming for large numbers of assets and observations, we approximate it on the S&P 1500 dataset via a cross-validation based on a minimum of 100 folds. Specifically, the number of folds is $n_{\text{cv}} = \lceil T_{\text{d}}/T_{\text{fold}} \rceil$, where T_{d} is the number of daily return observations and $T_{\text{fold}} = \lfloor T_{\text{d}}/100 \rfloor$ is the number of return observations in a fold. The estimator is then given as $\widehat{\Sigma}_{t,\text{CV}} = \frac{1}{T_{\text{d}}} \sum_{j=1}^{n_{\text{cv}}} \widehat{\Sigma}^{-1}_{t,-j} (R_{j,\text{CV}} - \bar{R}_{\text{CV}})(R_{j,\text{CV}} - \bar{R}_{\text{CV}})' \widehat{\Sigma}^{-1}_{t,-j}$.

$R_{b,t} = w'_{b,t}R_t$ and the return vector of characteristics-managed portfolios $R_{c,t} = X'_{t-1}R_t/N_t$. DeMiguel et al. (2013) show that mean-variance optimization for parametric portfolios leads to a tractable quadratic optimization problem with respect to θ. To deal with a large number of characteristics, DeMiguel et al. (2013) include a regularization constraint, i.e., an ℓ_1-norm constraint on the parameter vector θ. We follow these authors by adding an ℓ_1-norm penalty term with parameter λ and obtain the parameter vector of the regularized parametric portfolios as

$$\widehat{\theta}_t = \arg\min_{\theta} \frac{\gamma}{2}\theta'\widehat{\Sigma}_{c,t}\theta + \gamma\,\theta'\widehat{\sigma}_{bc,t} - \theta'\widehat{\mu}_{c,t} + \lambda\sum_{k=1}^{K}|\theta_k|, \tag{3.11}$$

where γ is the risk aversion parameter, $\widehat{\mu}_{c,t}$ the estimated expected return vector and $\widehat{\Sigma}_{c,t}$ the estimated covariance matrix of characteristics-managed portfolios. $\widehat{\sigma}_{bc,t}$ denotes the vector of estimated covariances of characteristics-managed portfolio returns with the benchmark portfolio return. Setting $\lambda = 0$, we obtain the unregularized mean-variance parametric portfolios for risk aversion γ. Letting $\lambda \to \infty$, we obtain the benchmark portfolio, which is independent from the risk aversion parameter. We determine the optimal regularization parameter λ empirically via the out-of-sample validation procedure described above which aims to maximize the out-of-sample Sharpe ratio. The covariance matrix $\widehat{\Sigma}_{c,t}$ is estimated via the shrinkage covariance matrix estimator in (3.10). The estimated parameter vector $\widehat{\theta}_t$ and the regularization parameter λ in (3.11) are updated at the end of each year based on characteristics and return data from the previous $T = 144$ months.

Let $\widehat{w}_t$ be any of the estimated portfolio weights for the N_t assets in the portfolio for month t in the out-of-sample testing period and R_t the vector of excess returns on the N_t firms from end of month $t-1$ to end of month t. The resulting realized out-of-sample portfolio excess return for month t is $R_{\text{PF},t} = \widehat{w}_t'R_t$. We evaluate the mean-variance optimized portfolio strategies by calculating the following out-of-sample portfolio statistics:

$$\widehat{\mu}_{\text{oos}} = \frac{1}{\mathcal{T}}\sum_{t=t_1}^{t_{\mathcal{T}}} R_{\text{PF},t}, \quad \widehat{\sigma}_{\text{oos}} = \sqrt{\frac{1}{\mathcal{T}-1}\sum_{t=t_1}^{t_{\mathcal{T}}}\left(R_{\text{PF},t} - \widehat{\mu}_{\text{oos}}\right)^2}, \quad \widehat{\text{SR}}_{\text{oos}} = \frac{\widehat{\mu}_{\text{oos}}}{\widehat{\sigma}_{\text{oos}}},$$

i.e., the out-of-sample mean return, standard deviation and Sharpe ratio of the portfolio, where $\mathcal{T}$ denotes the length, t_1 the first and $t_{\mathcal{T}}$ the last month in the out-of-sample period.

An important statistic concerning the practical implementation of portfolio strategies is the month by month portfolio turnover. We define the turnover associated with portfolio strategy

$\widehat{w}_t$ in month t as

$$\tau_t = \sum_{i=1}^{\widetilde{N}_t} |\widehat{w}_{t,i}^{-\text{in}} - \widehat{w}_{(t-1)^+,i}^{-\text{out}}| + \sum_{i=1}^{N_{t-1}^{\text{out}}} |\widehat{w}_{(t-1)^+,i}^{\text{out}}| + \sum_{i=1}^{N_t^{\text{in}}} |\widehat{w}_{t,i}^{\text{in}}|, \tag{3.12}$$

$$\text{where} \quad \widehat{w}_{t^+,i} = \widehat{w}_{t,i} \frac{1 + R_{f,t} + R_{t,i}}{1 + R_{f,t} + \widehat{w}_t' R_t}.$$

The superscripts "−in" and "−out" on portfolio weights $\widehat{w}$ refer to weight vectors without added or removed index constituents, the superscripts "in" and "out" on portfolio weights indicate that the weight vectors consist of only the added or removed portfolio constituents, respectively. $\widetilde{N}_t$ denotes the number of remaining portfolio constituents from $t-1$ to t, N_{t-1}^{out} the number of removed constituents from $t-1$ to t and N_t^{in} the number of added constituents from $t-1$ to t. $\widehat{w}_{t^+,i}$ is the portfolio weight for asset i at time t immediately before rebalancing of the portfolio, where $R_{f,t}$ denotes the risk-free rate of return in month t. The three summands for τ_t in (3.12) represent the turnover associated with firms remaining in the portfolio from month $t-1$ to t, firms leaving the portfolio to month t and firms joining the portfolio, respectively.

3.5.2 Empirical results

We begin with a statistical evaluation of predictions of the different expected return models on the three datasets based on the S&P 500, S&P 1500 and STOXX Europe 600 indices. Table 3.2 reports time-series averages of cross-sectional statistics of expected return predictions from the models from Section 3.4.1, i.e., linear regression (LR), random forest (RF), gradient boosted regression trees (XGB), artificial neural network (NN) and an ensemble (ENS) of linear regression, boosted regression trees and neural network.

The reported statistics are average, standard deviation, 10th as well as 90th percentiles of expected return predictions to summarize their distribution and variability. Average predictions are similar for all compared methods and amount to roughly one percent per month on all three datasets. Predicted expected returns are about 0.20 percentage points per month higher on the S&P 1500 than on the S&P 500 dataset which can be explained by greater expected returns on mid and small cap firms in the S&P 1500 compared to large cap firms in the S&P 500. The on average 0.10 percentage points lower estimated expected returns for European firms compared to the firms in the S&P 500 are due to our backwards-looking estimation procedure and the inferior average performance of European firms over the past decades. More important for portfolio optimization is the cross-sectional variability of expected return predictions. Cross-sectional standard deviations are very similar on the S&P 500 and STOXX Europe 600 datasets, as both contain a comparable number of firms. The random forest captures substantial variation in expected return predictions but displays the lowest cross-sectional standard deviations on all three datasets (0.53 on the S&P 500, 0.72 on the S&P 1500 and 0.52 on the STOXX Europe 600 datasets). In contrast to the S&P 500, the S&P 1500 includes smaller firms which results in wider-spread predictions especially on the upper end of the expected return spectrum which is, e.g., evidenced by similar 10th percentiles of 0.13 for the ensemble model's predictions and a larger 90th percentile of 2.09 on the S&P 1500 dataset compared to 1.67 on the S&P 500 dataset. Gradient boosted trees and the neural network model display the largest variation in expected return predictions. Standard deviations between 0.66 and 0.91 are roughly as large as average expected return predictions.

To evaluate these predictions statistically we report root mean squared prediction errors (RMSE). RMSE is generally lowest for the random forest and the ensemble model. Both models decrease prediction errors of machine learning methods through forecast averaging. Because minimizing RMSE is not our objective for determining the hyperparameters of the models, the comparison of models based on the RMSE is not overly important for the performance of mean-variance optimized portfolios. Gu et al. (2020) demonstrate that neural networks are able to achieve lower prediction errors than linear regression models and random forests when models are specifically tuned for this task. To obtain more insight into prediction accuracy, we regress predicted expected returns on realized returns, where both variables are cross-sectionally de-meaned each month within the out-of-sample period as suggested in Lewellen (2015). β and R^2 denote the slope coefficient and coefficient of determination of the estimated regression. The slope coefficient quantifies how well the dispersion of predictions aligns with true expected returns. A slope coefficient smaller than one indicates that on average predictions vary more than true expected returns and need to be shrunk by approximately $(1-\beta)$. Table 3.2 shows that slope coefficients are consistently smaller than one and are mostly larger than 0.3 which shows that estimates have predictive power for expected returns but are too widely dispersed on average. Predictions of the random forest exhibit the largest slope coefficients on all datasets but need to be shrunk by about 37 percent on average on the US datasets and by 67 percent on the European dataset. One reason for the low slope coefficients is that LR, XGB and NN could be regularized stronger for minimization of mean squared prediction errors. This would lead to predictions which are less varying, but as we aim to maximize out-of-sample portfolio Sharpe ratios this is not our objective. Another reason for the larger dispersion are estimation errors which the random forest as a tree ensemble method aims to reduce by averaging over predictions of individual trees. This benefit of forecast averaging is also observed for the ensemble model of linear regression, boosted trees and neural network. ENS consistently displays the second highest slope coefficients. R^2 coefficients show that when predictions are shrunk to match

Table 3.2: Statistics of expected return predictions from machine learning methods using firm characteristics for firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

	S&P 500					S&P 1500					STOXX Europe 600				
	LR	RF	XGB	NN	ENS	LR	RF	XGB	NN	ENS	LR	RF	XGB	NN	ENS
Expected returns															
Avg	0.86	0.86	0.89	0.87	0.88	1.07	1.07	1.08	1.07	1.07	0.78	0.78	0.78	0.79	0.78
Std	0.64	0.53	0.67	0.66	0.61	0.75	0.72	0.91	0.88	0.78	0.63	0.52	0.68	0.74	0.64
P10	0.09	0.22	0.07	0.09	0.13	0.16	0.27	−0.02	0.04	0.13	0.00	0.14	−0.09	−0.10	−0.01
P90	1.69	1.53	1.75	1.68	1.67	2.04	1.98	2.23	2.16	2.09	1.58	1.45	1.65	1.71	1.60
RMSE	9.79	9.78	9.79	9.79	9.79	12.27	12.26	12.27	12.27	12.26	9.42	9.42	9.42	9.42	9.42
β	0.40	0.62	0.46	0.45	0.50	0.48	0.63	0.52	0.51	0.59	0.28	0.33	0.31	0.26	0.31
R^2	0.09	0.15	0.13	0.13	0.13	0.12	0.18	0.19	0.17	0.19	0.05	0.05	0.07	0.06	0.07
Long-short portfolio															
Mean	8.49	11.39	10.98	10.38	10.65	11.61	15.12	14.78	14.43	15.58	6.83	6.82	8.14	7.53	8.16
Std	17.16	17.21	16.76	18.14	18.05	17.12	17.46	15.70	16.20	16.56	14.67	14.86	12.04	14.07	13.70
SR	0.49	0.66	0.66	0.57	0.59	0.68	0.87	0.94	0.89	0.94	0.47	0.46	0.68	0.54	0.60

Note: All numbers except for β and Sharpe ratio (SR) are in percent. The table reports time-series averages of the cross-sectional average, standard deviation, 10th and 90th percentiles as well as the root mean squared prediction error of monthly expected return predictions. β is the slope coefficient and R^2 the coefficient of determination of a pooled linear regression of predicted expected returns on realized returns where the explanatory variable as well as the response variable are cross-sectionally de-meaned on a month-by-month basis. LR, RF, XGB and NN denote the linear regression, random forest, gradient boosted regression trees and neural network models from Sections 3.4.1 and 3.4.3. ENS is an ensemble model averaging the predictions from LR, XGB and NN. Long-short portfolios are equally-weighted portfolios with long positions in stocks in the top quintile and short positions in stocks in the bottom quintile of expected return predictions. The table reports annualized means, standard deviations and Sharpe ratios of realized out-of-sample portfolio returns. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

the dispersion of expected returns, the predictions of nonlinear models explain more of the cross-sectional variation of expected returns compared to the linear model. ENS displays 33 to 61 percent higher R^2 coefficients than the linear regression model.

Table 3.2 exhibits out-of-sample performance statistics of long-short portfolios that we construct based on expected return predictions. Each month within the out-of-sample period we sort estimated expected returns into quintiles and consider a month-long equally-weighted long position in firms of the top decile and a short position in firms of the bottom decile. The resulting performance figures indicate the superiority of expected return predictions from machine learning methods. Mean realized out-of-sample returns of quintile spread portfolios are 20 to 34 percent larger for the ensemble model than for the linear model. For instance, on the S&P 1500 dataset the annualized mean portfolio return for ENS is 15.6 compared to 11.6 percent for LR over the 1995 to 2020 period. Out-of-sample Sharpe ratios of ENS improve by 19 to 39 percent upon the linear regression model obtaining an annualized Sharpe ratio of 0.94 for ENS on the S&P 1500 dataset.

In contrast to mean-variance optimized portfolios, long-short quintile or decile spread portfolios do not take portfolio risk, i.e., the information contained in the covariance matrix of individual stock returns, into account. They can thus be considered to be inefficient portfolios (Daniel et al. (2020)). We estimate mean-variance efficient portfolios as described in Section 3.4 for each month in the out-of-sample period and calculate averages and standard deviations of monthly standardized portfolio weights for stocks in quintile portfolios. The average mean-variance optimized portfolio weight for stocks in the bottom (top) quintile, i.e., the quintile of stocks with the lowest (highest) predicted expected returns, is about one standard deviation smaller (larger) than the average optimized portfolio weight. Portfolio weights in quintile portfolios two and four deviate by approximately −0.4 and 0.4 standard deviations from the average mean-variance optimized portfolio weight. These numbers illustrate that, as one would expect, higher predicted expected returns result in larger mean-variance optimized portfolio weights on average. But standard deviations of more than 0.6 quantify large variation of optimized portfolio weights within quintile portfolios implying that mean-variance optimized portfolios, that consider covariance matrix information, differ significantly from quintile portfolios that are based on expected returns. In many cases stocks within the lowest or highest quintiles of predicted expected returns are not among stocks with the most extreme portfolio weights. Average correlations of predicted expected returns with the mean-variance optimized portfolio weights are about 73 percent. The results hold for all three datasets and all considered prediction models and are summarized in Table 3.A.3 in the appendix.

Table 3.3 reports statistics on the out-of-sample performance of mean-variance optimized portfolios from Section 3.4.2 based on expected return predictions from sample means (SMV), the linear regression model, tree-based models and the neural network. We set the volatility target for optimized portfolios to $\sigma = 15\%$ per annum which should be comparable to the volatility of the equally-weighted market portfolio (EW). Other risk limits that we have tried result in equivalent out-of-sample Sharpe ratios because σ is a constant scaling factor of portfolio weights in (3.8). The scaling factor does not affect the relative relations of portfolio weights between assets which are determined by $\widehat{\Sigma}_t^{-1}\widehat{\mu}_t$. Considered benchmark portfolios are minimum variance portfolios estimated from 12 years of monthly data (MINV) and 24 months of daily data (MINVD) as well as regularized parametric portfolio policies (PPP). Table 3.3 reports in percent the annualized mean and standard deviation of realized out-of-sample returns for each portfolio strategy. Our primary evaluation criterion is the annualized Sharpe ratio on the S&P 500, S&P 1500 and STOXX Europe 600 datasets.

Table 3.3 demonstrates that using firm characteristics for estimating mean-variance optimal portfolios improves the out-of-sample performance of optimized portfolios considerably. Consistent with the literature, portfolios based on sample means (SMV) underperform equally-weighted portfolios in large asset spaces. Our main result is that mean-variance optimized portfolios based on expected return predictions from linear and nonlinear models obtain out-of-sample Sharpe ratios between 2.09 and 2.92 on the three datasets. These Sharpe ratios are more than three, four and five times larger than the Sharpe ratios of equally-weighted portfolios on the S&P 500, S&P 1500 and STOXX Europe 600 datasets, respectively. The mean-variance optimized portfolios, e.g., based on expected return predictions from the ensemble model ENS, realize higher out-of-sample Sharpe ratios than regularized parametric portfolios by 63% on the S&P 500, 59% on the S&P 1500 and 104% on the STOXX Europe 600 datasets. As reported in Table 3.3 we obtain the overall best out-of-sample performance for PPP with $\gamma = 20$. $\gamma = 5$, for instance, led to 8–12% lower Sharpe ratios than $\gamma = 20$.

On all three datasets the linear regression model delivers a competitive out-of-sample performance. The best performing model is an ensemble which averages the forecasts of the linear model, gradient boosted trees and neural network. Its Sharpe ratios of 2.35, 2.92 and 2.32 are 8–10% higher than the Sharpe ratios of the linear model. The neural network model displays slightly higher Sharpe ratios than LR by five to seven percent on the S&P 500, S&P 1500 and STOXX Europe 600 datasets, respectively. Employing the random forest model yields no improvements over the linear model in terms of Sharpe ratios of mean-variance optimized portfolios on our datasets. Gradient boosted trees are able to outperform LR only on the largest dataset obtaining a six percent higher out-of-sample Sharpe ratio. The results highlight that the larger advantages of using nonlinear models for constructing long-short quintile spread portfolios from expected return predictions as in Table 3.2 do not necessarily transfer to mean-variance optimized portfolios. By taking the covariance matrix of stock returns into account overall Sharpe

Table 3.3: Portfolio performance statistics for mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
Mean	10.72	7.60	6.69	−0.22	32.32	33.50	32.14	33.71	34.94	34.06
Std	16.44	11.21	11.39	16.25	15.06	16.03	14.80	14.97	14.85	23.65
Sharpe Ratio	0.65	0.68	0.59	−0.01	2.15	2.09	2.17	2.25	2.35	1.44
S&P 1500										
Mean	11.87	9.06	6.87	0.10	45.96	48.53	50.56	49.12	51.57	57.17
Std	18.88	10.81	10.46	16.81	17.28	18.61	17.88	17.65	17.68	31.11
Sharpe Ratio	0.63	0.84	0.66	0.01	2.66	2.61	2.83	2.78	2.92	1.84
STOXX Europe 600										
Mean	7.02	6.84	8.29	4.41	33.64	34.74	32.75	36.00	36.46	24.56
Std	17.03	10.16	9.62	18.04	15.69	16.50	15.60	15.71	15.70	21.55
Sharpe Ratio	0.41	0.67	0.86	0.24	2.14	2.11	2.10	2.29	2.32	1.14

Note: The table reports annualized means, standard deviations (both in percent) and Sharpe ratios of realized portfolio returns for ten portfolio strategies. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

ratios rise markedly but relative differences between the linear and nonlinear models diminish on our datasets. In Table 3.3 we still observe small benefits of employing nonlinear methods, e.g., artificial neural networks, to estimate expected returns for mean-variance optimized portfolios.

The optimized portfolios based on (3.8), i.e., SMV, LR, RF, XGB, NN and ENS, target an annualized out-of-sample volatility of 15%. Empirically, the volatilities of LR, XGB, NN and ENS are within one percentage point of the targeted volatility on the two smaller datasets. On the S&P 1500 dataset out-of-sample volatilities are just under 18% per annum. Overall the jackknife method does well to bring out-of-sample volatilities close to their target and deal with estimation risks in portfolio weights as described in Section 3.4.2. Observed out-of-sample volatilities are consistently lower than the volatility of the equally-weighted market portfolio. There remains some room for improvement, e.g., via an extension of the method for time-varying second moments as in Basak et al. (2009), which is beyond the scope of this paper. Out-of-sample volatilities of parametric portfolio policies, however, are not comparable because they are not constructed to meet the specific volatility target.

Naturally, minimum variance portfolios display the lowest out-of-sample volatilities. Their volatilities are 30 to 45 percent lower than the volatilities of the equally-weighted market portfolio. Using daily return data for covariance matrix estimation yields an improvement for minimum variance portfolios on the S&P 1500 and STOXX Europe 600 datasets, that is, MINVD obtains three and five percent lower volatilities than MINV. In terms of out-of-sample Sharpe ratios significant improvements over the equally-weighted portfolio can be observed for MINV on the S&P 1500 and STOXX Europe 600 and for MINVD only on the STOXX Europe 600 dataset. Nevertheless, we find that estimated mean-variance efficient portfolios benefit greatly in terms of out-of-sample Sharpe ratios from the use of daily data for covariance matrix estimation. On the three datasets realized Sharpe ratios for the linear model are 35, 43 and 61 percent higher, respectively, when compared to using monthly return data for estimation of the covariance matrix. Table 3.A.4 in the appendix summarizes the results when estimating the covariance matrix for optimized portfolios based on monthly data where we often only have about 60 monthly observations available for estimation. In particular, annualized out-of-sample volatilities of characteristics-based optimized portfolios are between 12.7 and 14.5 percent on all datasets and thereby below the target of 15 percent per annum. The jackknife procedure appears to benefit from monthly data or otherwise might benefit from time-varying covariance matrices for daily data.

The distributions of realized out-of-sample returns of mean-variance optimized portfolios are remarkably close to normal. On the three datasets the skewness coefficients for ENS are between -0.2 and 0.1 and the kurtosis coefficients between 2.9 and 4.0, while for the equally-weighted market portfolio skewness is about -0.5 and kurtosis is higher than 5.0. Realized returns from minimum variance portfolios display even more negative skewness coefficients which are about -1.0 on the US datasets with a kurtosis that is larger than 5.0. On the other hand, returns on parametric portfolios show large positive skewness of 0.9 and 1.9 on the US datasets but also large kurtosis of 7.4 and 13.7, see Table 3.A.5 in the appendix.

Table 3.4 presents annualized alphas (in percent) as well as t statistics of alphas of mean-variance optimized portfolios over the out-of-sample periods. Alphas are calculated with respect to the Fama-French six-factor model containing market, size, value, profitability, investment and momentum factors. To be able to compare alphas consistently across portfolio strategies we scale portfolio returns such that their out-of-sample standard deviations are equal to the standard deviation of the market factor. While the equally-weighted and minimum variance portfolios exhibit insignificant alphas on the US datasets, our optimized portfolios generate large positive mean returns that cannot be explained by exposures to the six factors. t statistics, calculated using Newey-West standard errors, are above 8 implying that alphas are statistically significant.

Table 3.4: Alphas of mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices with respect to the Fama-French six-factor model.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
Alpha	0.38	1.26	0.99	−0.90	28.87	28.24	29.76	31.53	32.32	13.89
t statistic	0.47	0.49	0.35	−0.35	7.96	8.07	7.97	8.52	8.39	4.48
S&P 1500										
Alpha	0.29	1.53	1.09	−1.47	37.45	37.03	40.50	40.27	42.09	21.14
t statistic	0.41	0.64	0.37	−0.45	9.92	9.20	10.55	10.24	10.22	4.84
STOXX Europe 600										
Alpha	3.07	4.63	8.68	−0.48	30.49	28.67	29.94	32.78	33.18	12.16
t statistic	5.14	1.72	2.77	−0.16	8.86	9.53	9.64	10.48	10.27	4.06

Note: The table reports annualized alphas (in percent) with respect to the Fama-French six-factor model as well as *t* statistics of alphas (based on Newey-West standard errors) of realized portfolio returns for ten portfolio strategies. Portfolio returns are scaled such that they have the same out-of-sample standard deviation as the market factor. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The data on the six-factor model are from Kenneth French's website, we convert returns on the six European factors from USD to EUR to be consistent within the analysis. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

The portfolio strategy based on the ensemble model, ENS, generates the largest annualized alphas amounting to 32 and 42 percent on the US datasets and 33 percent on the European data. Our mean-variance optimized portfolios outperform regularized parametric portfolio policies whose annualized alphas are 14, 21 and 12 percent, respectively. We thus obtain an increase in out-of-sample alpha coefficients of 130, 100 and 170 percent when employing our mean-variance optimized portfolio strategies from Section 3.4.

Figure 3.2 plots realized cumulative log excess returns of the estimated mean-variance optimal portfolio strategies over the out-of-sample period of each dataset. ENS confirms its position as the overall best strategy as it is consistently among the best performing models on the three datasets. Its outperformance of the equally-weighted portfolio is pervasive over the entire out-of-sample periods and it appears less affected by the large market drops due to crises in 2008 and 2020. The performance of optimized portfolios has degraded since the late 2000's and the financial crisis on the US datasets.

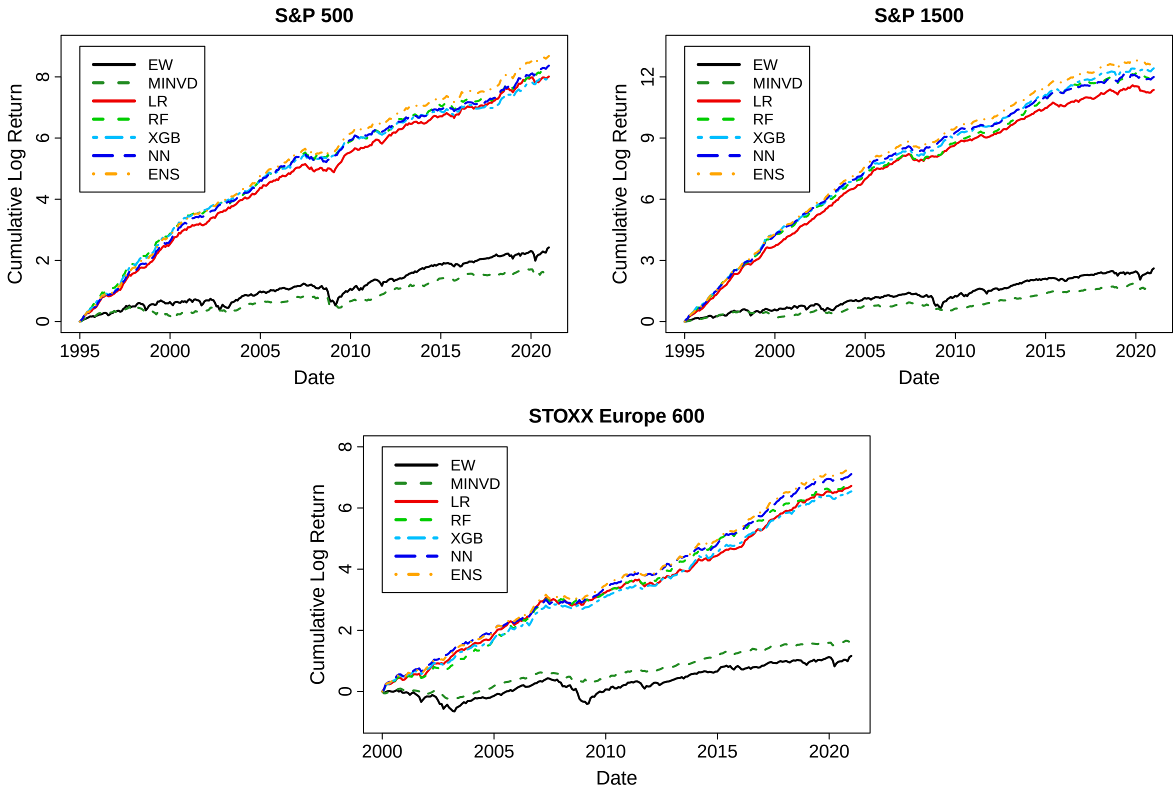

Figure 3.2: Out-of-sample cumulative logarithmic excess returns of portfolios based on firms from the S&P 500, S&P 1500 and STOXX Europe 600 indices. Out-of-sample periods start in 01/1995 for the S&P indices and in 01/2000 for the STOXX index and end in 12/2020. EW and MINVD are an equally-weighted portfolio and an estimated minimum variance portfolio based on 24 months of daily returns via the shrinkage covariance matrix estimator in (3.10). LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics.

Table 3.5 presents annualized performance figures for all considered strategies for two subperiods. The realized Sharpe ratios of the equally-weighted portfolios are roughly constant on the first subperiod until 12/2007 and the second subperiod starting in 01/2008. In contrast, Sharpe ratios of estimated optimal portfolio weights based on firm characteristics decrease by 40 to 50 percent on the US datasets when comparing the period after 2008 to before 2008. For instance, ENS attains a Sharpe ratio of 4.01 over the 1995 to 2007 period on the S&P 1500 dataset, while over the 2008 to 2020 period its Sharpe ratio is 2.05. This, however, still implies a 3.5 times outperformance of the equally-weighted market portfolio over the latter subperiod. Parametric portfolio policies suffer from the same phenomenon as its out-of-sample Sharpe ratios more than halve after 2008. On the STOXX Europe 600 dataset no such drop in performance

Table 3.5: Portfolio performance statistics on two subperiods for mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
01/1995 – 12/2007										
Mean	9.97	5.81	6.81	0.14	39.29	42.63	41.99	42.14	43.28	54.50
Std	14.56	10.55	10.40	17.12	13.92	15.75	14.96	14.19	14.32	28.62
Sharpe Ratio	0.68	0.55	0.65	0.01	2.82	2.71	2.81	2.97	3.02	1.90
01/2008 – 12/2020										
Mean	11.48	9.39	6.57	−0.57	25.34	24.36	22.29	25.28	26.60	13.63
Std	18.17	11.84	12.33	15.39	15.91	15.92	14.13	15.37	15.02	15.27
Sharpe Ratio	0.63	0.79	0.53	−0.04	1.59	1.53	1.58	1.65	1.77	0.89
S&P 1500										
01/1995 – 12/2007										
Mean	11.28	8.03	7.19	4.36	63.25	63.75	65.71	67.14	68.74	99.38
Std	15.80	9.90	9.26	17.31	16.47	18.87	18.09	16.90	17.16	38.09
Sharpe Ratio	0.71	0.81	0.78	0.25	3.84	3.38	3.63	3.97	4.01	2.61
01/2008 – 12/2020										
Mean	12.46	10.09	6.54	−4.16	28.68	33.31	35.44	30.63	34.41	14.96
Std	21.58	11.68	11.57	16.25	16.67	17.32	16.59	16.87	16.82	13.88
Sharpe Ratio	0.58	0.86	0.57	−0.26	1.72	1.92	2.14	1.82	2.05	1.08
STOXX Europe 600										
01/2000 – 12/2007										
Mean	5.05	5.81	7.76	2.92	38.46	39.98	37.18	39.04	40.70	33.20
Std	16.51	9.59	9.26	18.09	16.82	17.76	17.39	16.54	17.20	27.80
Sharpe Ratio	0.31	0.61	0.84	0.16	2.29	2.25	2.14	2.36	2.37	1.19
01/2008 – 12/2020										
Mean	8.24	7.47	8.62	5.33	30.68	31.52	30.02	34.14	33.86	19.23
Std	17.38	10.52	9.86	18.06	14.95	15.66	14.39	15.20	14.70	16.49
Sharpe Ratio	0.47	0.71	0.87	0.30	2.05	2.01	2.09	2.25	2.30	1.17

Note: The table reports annualized means, standard deviations (both in percent) and Sharpe ratios of realized portfolio returns for ten portfolio strategies. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

of optimized portfolios can be observed. Sharpe ratios of estimated mean-variance efficient portfolios are about constant over the 1995 to 2007 and 2008 to 2020 periods. ENS attains Sharpe ratios of 2.37 and 2.30 over the first and second subperiods, respectively, outperforming the equally-weighted portfolio and parametric portfolios by large margins.

To assess the similarity of portfolio weights from different strategies we calculate time-series averages of monthly correlation matrices for portfolio weights of MINVD, LR, RF, XGB, NN, ENS and PPP as given in Table 3.A.6 in the appendix. Minimum variance portfolio weights exhibit a correlation of no more than 0.31 with the mean-variance optimized portfolio weights highlighting the influence of estimated expected returns on portfolio weights. Correlations of portfolio weights for mean-variance optimized portfolios LR, RF, XGB and NN are between 0.64 and 0.84 over the three datasets. Excluding the correlations of LR, XGB, NN with ENS which are larger by construction of ENS, the pair of linear model and neural network and the pair of tree-based methods (random forest and gradient boosted trees) display the highest correlations of portfolio weights with about 0.80 on average. Parametric portfolio weights are moderately correlated with the mean-variance optimized portfolio weights based on LR, RF, XGB, NN and ENS exhibiting correlations between 0.34 and 0.58.

Table 3.6 provides insights on the properties of estimated mean-variance efficient portfolios. The table reports statistics on portfolio weights such as the sum of weights in risky assets, the average absolute weight, the maximum and minimum weights as well as the fraction of negative weights and average monthly portfolio turnover. Portfolios allowing for an investment into the risk-free asset, i.e., SMV, LR, RF, XGB, NN and ENS, are on average 41 to 83 percent invested in risky assets. The fraction invested into risky assets has to do with the specified volatility target, a lower σ would imply less capital invested into risky assets. The gradient boosted trees and the neural network portfolios having the lowest average investment into risky assets appear to tilt investments towards higher volatility stocks. While their average sum of weights ranges from 41 to 66 percent over the three datasets, the random forest model displays the highest average investment into risky assets ranging from 54 to 82 percent.

The fraction of negative portfolio weights is between 47 and 50 percent for all machine learning portfolios, that is, nearly half of the positions in risky assets are short positions. The fraction of short positions is around 40 percent for PPP and minimum variance portfolios. These numbers suggest that short positions are relevant for and short-sale constraints could be detrimental to portfolio performance as evidenced in Brandt et al. (2009). However, in practice the estimated mean-variance efficient portfolios could be used to deviate from an existing position in a respective market index. By this means an outperformance of the market index can be achieved without engaging in short positions at the overall portfolio level. As argued in Ledoit and Wolf (2017) this can be useful for portfolio managers that are benchmarked against a certain

Table 3.6: Portfolio weight statistics for mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
Sum Weights	100.00	100.00	100.00	49.52	49.54	54.53	42.10	44.30	47.84	100.00
Avg Abs Weight	0.28	0.65	1.18	2.33	2.60	2.59	2.69	2.56	2.59	1.18
Max Weight	0.28	2.39	5.47	13.32	12.13	11.95	11.43	10.85	11.18	5.81
Min Weight	0.28	−2.42	−4.18	−8.39	−10.18	−10.23	−10.89	−10.91	−10.17	−4.60
Frac Neg Weights	0.00	32.81	43.07	50.38	49.15	48.95	49.11	48.29	48.89	38.54
Turnover	7.31	40.99	124.78	243.83	527.17	672.72	740.97	628.38	591.56	165.21
S&P 1500										
Sum Weights	100.00	100.00	100.00	54.43	78.50	81.62	65.56	67.12	72.37	100.00
Avg Abs Weight	0.10	0.22	0.40	0.84	0.97	0.96	1.05	0.99	0.99	0.62
Max Weight	0.10	0.86	1.95	14.30	4.62	5.01	4.79	4.30	4.34	4.21
Min Weight	0.10	−0.95	−1.50	−3.68	−4.12	−4.18	−4.56	−4.63	−4.31	−2.79
Frac Neg Weights	0.00	30.34	40.76	50.42	47.93	47.73	47.97	47.51	47.86	42.51
Turnover	8.94	35.97	112.33	252.29	503.15	690.57	761.22	645.35	627.07	305.04
STOXX Europe 600										
Sum Weights	100.00	100.00	100.00	82.28	67.51	80.94	59.56	60.12	65.99	100.00
Avg Abs Weight	0.26	0.61	0.88	2.21	2.32	2.31	2.47	2.35	2.35	1.13
Max Weight	0.26	2.50	6.41	12.52	9.86	10.32	10.37	10.20	9.77	4.68
Min Weight	0.26	−2.10	−2.83	−8.34	−9.29	−8.63	−9.72	−9.54	−9.08	−5.16
Frac Neg Weights	0.00	34.20	41.93	48.55	47.93	47.59	47.77	47.99	47.82	40.91
Turnover	7.83	42.46	95.37	234.39	537.70	647.10	726.84	675.09	610.82	201.56

Note: All numbers are in percent. The table reports monthly averages of the sum of weights, the average absolute weight, the maximum and minimum weights, the fraction of negative weights and portfolio turnover based on (3.12) for ten portfolio strategies. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

index and whose actively managed part of the portfolio can thus be considered separately from the passive benchmark weights. In this comparison, we consider the best potential performance of each portfolio strategy. As imposing short-sale constraints would affect each strategy, the preferred unconstrained portfolio strategy would still be advantageous.

Figure 3.3 plots monthly portfolio turnover for estimated minimum variance and mean-variance optimal portfolios based on LR, RF, XGB, NN and ENS over the out-of-sample periods. Portfolio turnover is notably high for the mean-variance optimized portfolios, especially those

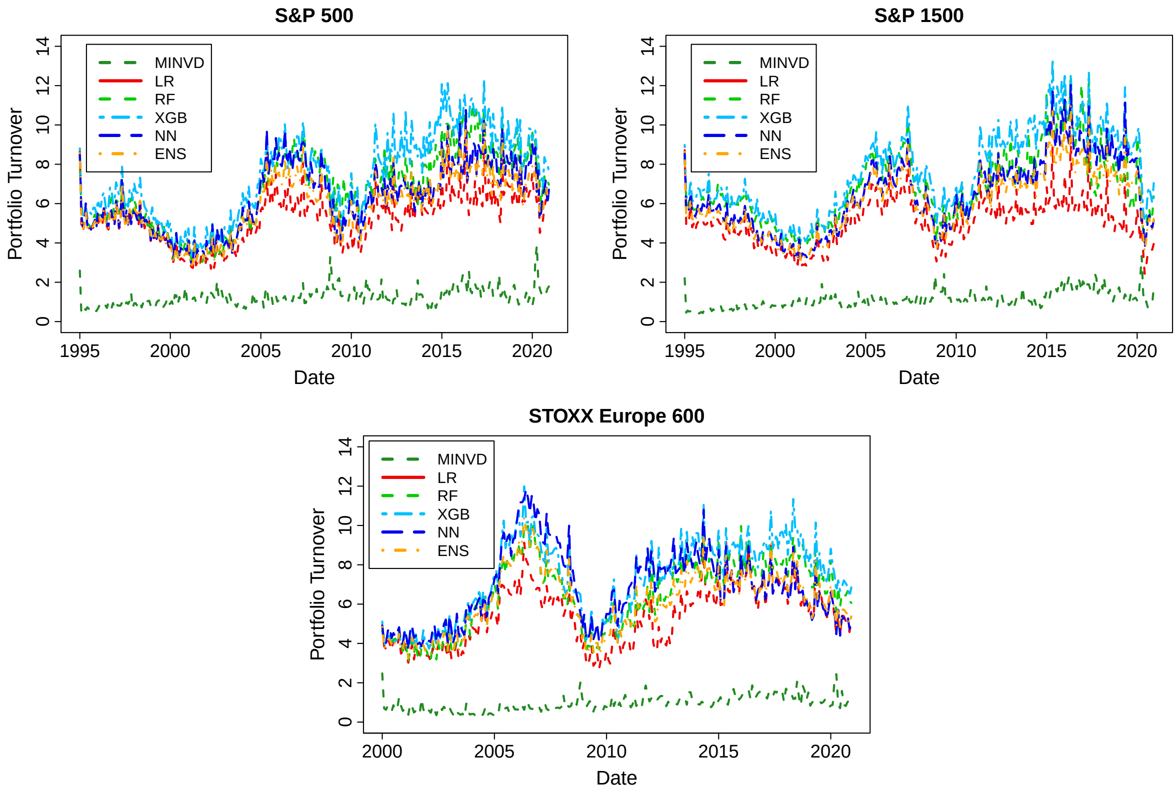

Figure 3.3: Monthly turnover of portfolios based on firms from the S&P 500, S&P 1500 and STOXX Europe 600 indices. Out-of-sample periods start in 01/1995 for the S&P indices and in 01/2000 for the STOXX index and end in 12/2020. MINVD is an estimated minimum variance portfolio based on 24 months of daily returns via the shrinkage covariance matrix estimator in (3.10). LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics.

based on nonlinear methods. As summarized in Table 3.6, average monthly turnover exceeds 600 percent for the random forest, boosted regression trees and neural network model on all three datasets. The average absolute, maximum and minimum portfolio weights show that the XGB and NN portfolios result in more dispersed portfolio weights with particularly larger short positions compared to LR and RF contributing to their higher turnover. LR portfolios produce consistently less amounts of monthly turnover still exceeding 500 percent on average. Parametric portfolio policies are advantageous in this respect with average monthly turnovers between 165 and 305 percent despite their higher out-of-sample portfolio volatility. Naturally, portfolio turnover is lowest for the equally-weighted portfolios whose weights are not affected by estimation risks and time-variation in return moments, i.e., expected returns and covariance matrix. Average monthly turnover for the equally-weighted portfolio is highest with 9 percent on the S&P 1500 dataset. Minimum variance portfolios based on monthly data produce an average turnover of 40 percent as opposed to 95 to 125 percent for the minimum variance portfolios based on daily data which can be explained by their shorter estimation windows. In the following section we attempt to reduce the excessive turnover of the mean-variance optimized portfolios over time, especially for those portfolios based on the nonlinear machine learning methods.

3.5.3 Turnover-reduced portfolios

In order to reduce the enormous amounts of turnover of the optimized portfolios from Section 3.5.2 we explore an approach of combining month t portfolio weights with the weights from the previous period such that a fixed amount of turnover is not exceeded. This convex combination of portfolio weights is similar to Füss et al. (2014), but we target a fixed amount of turnover whereas Füss et al. (2014) minimize the out-of-sample portfolio variance. Let $\widetilde{w}_t$ denote portfolio weights for month t of a turnover-reduced strategy and let $\widetilde{w}_t' = (\widetilde{w}_t^{-\text{in},\prime}, \widetilde{w}_t^{\text{in},\prime})$ with superscripts defined analogously to (3.12). With κ_t as the shrinkage coefficient on month t portfolio weights we have

$$\widetilde{w}_t^{-\text{in}} = \kappa_t\, \widehat{w}_t^{-\text{in}} + (1-\kappa_t)\, \widetilde{w}_{(t-1)^+}^{-\text{out}} \quad \text{and} \quad \widetilde{w}_t^{\text{in}} = \kappa_t\, \widehat{w}_t^{\text{in}}. \tag{3.13}$$

The left equation in (3.13) shrinks newly estimated portfolio weights for month t towards portfolio weights from the previous period immediately before rebalancing takes place on the assets remaining in the portfolio from the previous period. The right equation in (3.13) shrinks newly estimated portfolio weights for month t to zero on the assets which join the portfolio for month t and thus do not have any portfolio weights from the previous period. The shrinkage approach implies that, if $\kappa_j < 1$ for $j = t_0, \ldots, t$, portfolio weights $\widetilde{w}_t$ of the turnover-reduced strategy depend on the entire history of optimized portfolio weights $\widehat{w}_j$ with decreasing weights on earlier periods.

Requiring portfolio turnover of $\widetilde{w}_t$ to be equal to a fixed value of η yields

$$\kappa_t = \frac{\eta - \sum_{i=1}^{N_{t-1}^{\text{out}}} |\widetilde{w}_{(t-1)^+,i}^{\text{out}}|}{\sum_{i=1}^{\widetilde{N}_t} |\widehat{w}_{t,i}^{-\text{in}} - \widetilde{w}_{(t-1)^+,i}^{-\text{out}}| + \sum_{i=1}^{N_t^{\text{in}}} |\widehat{w}_{t,i}^{\text{in}}|}. \tag{3.14}$$

from (3.12). Two special cases arise. First, when the turnover associated with firms leaving the portfolio already exceeds the prespecified value of η, κ_t turns negative. We then force κ_t to be zero resulting in a portfolio turnover higher than η. Thus, if η is chosen too small, the turnover constraint cannot always be satisfied. Second, when the turnover associated with rebalancing

Table 3.7: Portfolio performance statistics for turnover-reduced mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
$\eta = 1/6$										
Mean	10.72	7.99	7.25	3.65	17.79	18.02	15.90	16.38	17.50	20.60
Std	16.44	11.34	10.74	15.23	10.19	10.80	10.07	9.80	10.14	13.66
Sharpe Ratio	0.65	0.70	0.68	0.24	1.75	1.67	1.58	1.67	1.73	1.51
$\eta = 1/3$										
Mean	10.72	7.53	6.39	0.94	21.17	21.03	18.49	19.61	20.92	23.20
Std	16.44	11.32	11.16	15.52	10.23	10.65	9.78	9.74	10.09	15.09
Sharpe Ratio	0.65	0.67	0.57	0.06	2.07	1.97	1.89	2.01	2.07	1.54
$\eta = 1$										
Mean	10.72	7.58	6.14	0.08	26.39	26.71	24.02	25.46	27.09	29.16
Std	16.44	11.21	11.44	15.94	12.49	12.44	11.32	11.67	12.09	18.87
Sharpe Ratio	0.65	0.68	0.54	0.01	2.11	2.15	2.12	2.18	2.24	1.55
S&P 1500										
$\eta = 1/6$										
Mean	11.85	9.34	7.66	7.19	22.35	21.26	19.79	19.94	21.61	24.62
Std	18.86	11.07	10.39	13.98	12.35	11.91	10.98	10.78	11.46	13.01
Sharpe Ratio	0.63	0.84	0.74	0.51	1.81	1.78	1.80	1.85	1.89	1.89
$\eta = 1/3$										
Mean	11.87	9.19	6.84	4.84	26.49	25.17	23.63	24.06	26.07	28.78
Std	18.88	10.96	10.66	14.88	12.69	11.96	10.72	10.89	11.49	14.44
Sharpe Ratio	0.63	0.84	0.64	0.33	2.09	2.11	2.20	2.21	2.27	1.99
$\eta = 1$										
Mean	11.87	9.06	6.59	0.60	33.97	32.23	31.87	32.34	35.09	37.02
Std	18.88	10.81	10.60	16.61	14.22	13.31	11.77	12.23	12.81	19.19
Sharpe Ratio	0.63	0.84	0.62	0.04	2.39	2.42	2.71	2.64	2.74	1.93

Continued on next page

the portfolio is smaller than η, we set κ_t to one which implies that portfolio weights $\widetilde{w}_t$ are equal to the optimized portfolio weights $\widehat{w}_t$ for month t.

Table 3.7 summarizes portfolio performance statistics for three levels of η corresponding to portfolio turnovers of 100 percent every half year ($\eta = 1/6$), every quarter ($\eta = 1/3$) and every month ($\eta = 1$). We find that restricting monthly turnover to $\eta = 1/6$ leads to a substantial reduction of portfolio performance. Mean out-of-sample returns of mean-variance optimized portfolios often halve as a result of the turnover constraint and are between 10 and 23 percent per annum. The LR-based portfolios are just slightly less affected than portfolios based on machine learning methods. Interestingly, portfolio volatilities decrease as well as an effect of the turnover

Table 3.7 (continued).

Portfolio Statistics	EW	SMV	MINV	MINVD	LR	RF	XGB	NN	ENS	PPP
STOXX Europe 600										
$\eta = 1/6$										
Mean	7.02	7.58	8.10	2.62	12.79	11.36	10.34	11.31	12.23	12.82
Std	17.02	9.98	8.85	11.91	10.39	10.91	9.90	10.08	10.46	13.52
Sharpe Ratio	0.41	0.76	0.92	0.22	1.23	1.04	1.04	1.12	1.17	0.95
$\eta = 1/3$										
Mean	7.02	6.98	7.77	3.13	17.02	15.02	14.14	15.63	16.71	15.88
Std	17.03	10.23	9.35	14.27	11.10	11.64	10.57	10.57	11.12	14.44
Sharpe Ratio	0.41	0.68	0.83	0.22	1.53	1.29	1.34	1.48	1.50	1.10
$\eta = 1$										
Mean	7.02	6.79	7.87	3.91	24.20	22.42	21.09	23.56	24.77	21.40
Std	17.03	10.19	9.68	17.26	13.17	13.53	12.26	12.27	12.94	18.17
Sharpe Ratio	0.41	0.67	0.81	0.23	1.84	1.66	1.72	1.92	1.91	1.18

Note: The table reports annualized means, standard deviations (both in percent) and Sharpe ratios of realized portfolio returns for different levels η of tolerated monthly turnover. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the turnover-reduced mean-variance optimized portfolios as defined in (3.13) and (3.14) based on the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the turnover-reduced version of the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

constraint and are between 9.80 and 12.35 percent on all three datasets. Minimum variance portfolios estimated from daily data, i.e., MINVD, benefit with a reduction in out-of-sample volatility of six and eight percent on the S&P 500 and STOXX Europe 600 datasets outperforming MINV on all three datasets. Sharpe ratios of mean-variance optimized portfolios drop by 18 to 27 percent on the S&P 500, by 32 to 35 percent on the S&P 1500 and by 43 to 51 percent on the STOXX Europe 600 datasets. Again, LR-based portfolios are slightly less affected than portfolios based on machine learning methods. Nevertheless, the out-of-sample Sharpe ratios are overall about three times as large as those of the equally-weighted portfolios. In contrast, parametric portfolio policies are more robust with respect to the turnover constraint, they even benefit slightly in terms of Sharpe ratios on the two US datasets, and exhibit portfolio volatilities between 13 and 14 percent per annum.

Raising the permitted amount of monthly turnover increasingly recovers the performance of unconstrained portfolios. For instance, on the S&P 500 dataset Sharpe ratios of the turnover-reduced ENS portfolio are at 74, 88 and 95 percent of the Sharpe ratios of its unconstrained version for $\eta = 1/6,\ 1/3$, and 1. On the STOXX Europe 600 dataset the ratios are 50, 65 and 82 percent, respectively, illustrating the higher impact of the turnover constraint. We find that constraining portfolio turnover to 100 percent per month, which is at least a six-times reduction of turnover for the machine-learning-based portfolios, does not substantially hurt the out-of-sample portfolio performance of mean-variance optimized portfolios. The turnover-reduced NN and ENS portfolios outperform turnover-reduced LR portfolios on all three datasets in terms of Sharpe ratios. Assuming proportional round-trip transaction costs of 0.5% (as in DeMiguel et al. (2009)), we find that the turnover-reduced ENS portfolio with $\eta = 1$ obtains annualized Sharpe ratios net of transaction costs of 1.75, 2.27 and 1.45 on the S&P 500, S&P 1500 and STOXX Europe 600 datasets as opposed to 0.63, 0.60 and 0.39 for the equally-weighted portfolio, respectively. This is a gain of 180 percent (S&P 500) and 280 percent (S&P 1500, STOXX Europe 600) in out-of-sample net Sharpe ratios on the three datasets. Lower levels of transaction costs, which are realistic for investors in modern times, further increase the gain in net Sharpe ratio. Compared with turnover-reduced parametric portfolio policies based on $\eta = 1/3$, which implies a 100 percent portfolio turnover every quarter striking a favorable balance between portfolio performance and turnover for PPP, we obtain gains in net Sharpe ratios of at least 23, 22 and 51 percent, respectively.

Table 3.8 presents performance statistics for turnover-reduced portfolios with $\eta = 1$ on two subperiods. As in Table 3.5 mean-variance optimized portfolios attain superior Sharpe ratios on the US datasets over the 1995–2007 period compared to the 2008–2020 period, but these performance differences are smaller for turnover-reduced portfolios. The turnover constraint primarily decreases Sharpe ratios over the first subperiod. On the S&P 500 dataset Sharpe ratios of ENS net of transaction costs of 0.5% are 3.3 and 2.2 times the Sharpe ratio of the equally-weighted portfolio over the former and latter subperiods, respectively. On the S&P 1500 dataset net Sharpe ratios of ENS are 4.4 and 2.9 times the Sharpe ratio of the equally-weighted portfolio. Parametric portfolio weights with $\eta = 1/3$ are outperformed by 54 and 47 percent on the US datasets over the more recent subperiod.[4] On the STOXX Europe 600 dataset we obtain 5.7 and 3.1 times higher net Sharpe ratios than the equally-weighted portfolio over the first and second subperiods as well as 34 and 66 percent higher net Sharpe ratios compared to parametric portfolio policies with $\eta = 1/3$.

The analysis of turnover-constrained portfolios demonstrates that it is practically feasible to reduce portfolio turnover without sacrificing the outperformance of estimated mean-variance

[4] The results for turnover-reduced portfolio strategies with $\eta = 1/3$ over the two subperiods are not reported in order to save space but are consistent with the results in Table 3.7 and are available upon request.

Table 3.8: Portfolio performance statistics on two subperiods for turnover-reduced mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 with tolerated monthly turnover of $\eta = 1$.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
01/1995 – 12/2007										
Mean	9.97	5.81	6.52	0.17	30.49	32.74	29.96	30.49	32.04	44.44
Std	14.56	10.55	10.35	16.50	11.98	13.09	11.53	11.69	11.92	20.99
Sharpe Ratio	0.68	0.55	0.63	0.01	2.54	2.50	2.60	2.61	2.69	2.12
01/2008 – 12/2020										
Mean	11.48	9.36	5.76	−0.01	22.30	20.67	18.08	20.43	22.13	13.89
Std	18.17	11.84	12.47	15.40	12.91	11.54	10.87	11.50	12.12	15.31
Sharpe Ratio	0.63	0.79	0.46	0.00	1.73	1.79	1.66	1.78	1.83	0.91
S&P 1500										
01/1995 – 12/2007										
Mean	11.28	8.03	7.07	6.26	44.38	41.53	39.94	42.32	44.51	58.95
Std	15.80	9.90	9.24	17.57	13.50	13.24	12.42	12.34	12.89	21.33
Sharpe Ratio	0.71	0.81	0.77	0.36	3.29	3.14	3.22	3.43	3.45	2.76
01/2008 – 12/2020										
Mean	12.46	10.09	6.12	−5.07	23.56	22.94	23.80	22.37	25.66	15.08
Std	21.58	11.68	11.84	15.48	14.33	12.87	10.63	11.45	12.18	14.26
Sharpe Ratio	0.58	0.86	0.52	−0.33	1.64	1.78	2.24	1.95	2.11	1.06
STOXX Europe 600										
01/2000 – 12/2007										
Mean	5.05	5.77	7.76	4.29	28.91	28.54	24.63	27.63	28.97	28.31
Std	16.51	9.60	9.26	17.57	14.43	14.93	14.17	13.62	14.61	21.72
Sharpe Ratio	0.31	0.60	0.84	0.24	2.00	1.91	1.74	2.03	1.98	1.30
01/2008 – 12/2020										
Mean	8.24	7.41	7.94	3.68	21.30	18.65	18.90	21.06	22.18	17.15
Std	17.38	10.57	9.96	17.13	12.30	12.51	10.92	11.35	11.78	15.54
Sharpe Ratio	0.47	0.70	0.80	0.21	1.73	1.49	1.73	1.86	1.88	1.10

Note: The table reports annualized means, standard deviations (both in percent) and Sharpe ratios of realized portfolio returns for a tolerated monthly turnover of $\eta = 1$. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the turnover-reduced mean-variance optimized portfolios as defined in (3.13) and (3.14) based on the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the turnover-reduced version of the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

efficient portfolios with respect to the equally-weighted market portfolio or regularized parametric portfolio policies. We find that using machine learning methods such as artificial neural networks to estimate expected returns from firm characteristics can slightly improve out-of-sample performance of mean-variance optimized portfolios compared to a linear regression model. Mean-variance optimized portfolios are able to outperform parametric portfolio policies in terms of out-of-sample net Sharpe ratios especially over the more recent time period from 2008 to 2020. The economic value of our estimated efficient portfolios can also be emphasized by calculating the risk-free rate that needs to be added to the value-weighted market index return so that investors are indifferent with respect to the net Sharpe ratio between the two portfolio strategies. That is, for portfolio strategy J we calculate x such that $\left(\widehat{\mu}_{\text{oos}}^{\text{VW}} + x\right) / \widehat{\sigma}_{\text{oos}}^{\text{VW}} = \left(\widehat{\mu}_{\text{oos}}^{J} - \pi\tau_{\text{oos}}^{J}\right) / \widehat{\sigma}_{\text{oos}}^{J}$, where $\tau_{\text{oos}}^{\text{J}}$ denotes the turnover of portfolio strategy J over the out-of-sample period and π are the proportional round-trip transaction costs, in our case we set $\pi = 0.5\%$. The value of x is equivalent to the gain in certainty equivalent based on mean-variance utility from portfolio strategy J compared to the market index portfolio, where the portfolio weights of J are scaled such that its out-of-sample volatility is equal to the market volatility. For the turnover-reduced ENS portfolio with $\eta = 1$ we find that investors would require 17, 25 and 19 percent per annum in excess of the S&P 500, S&P 1500 and STOXX Europe 600 in order to be indifferent. For the optimized portfolio based on the linear model, LR, we find 16, 21 and 18 percent, respectively, and for parametric portfolio weights with $\eta = 1/3$ investors would require an additional 12 to 19 percent per annum in excess of the market index return to be indifferent. These results, that are summarized in Table 3.A.7 in the appendix, further stress the significant economic gain from our mean-variance optimized portfolios.

3.6 Conclusion

In this paper, we study the economic value of large mean-variance optimal portfolios making use of machine learning methods to deal with parameter estimation risks. We predict expected stock returns from monthly firm characteristics and employ higher frequency daily data in a shrinkage covariance matrix estimator based on the single-index model. In a comparative empirical analysis we evaluate the portfolios' out-of-sample performance for three investment universes based on major stock market indices, i.e., the S&P 500, S&P 1500 and STOXX Europe 600, which provide for natural benchmarks of US and European equity markets and ensure that we solely invest in liquid stocks. In the empirical evaluation, our estimated efficient portfolios exhibit annualized Sharpe ratios of 2.35, 2.92 and 2.32 between 1995 to 2020 for the US indices and between 2000 to 2020 for the European index. The gains in out-of-sample Sharpe ratios are 260, 360 and

460 percent compared to equally-weighted portfolios over the same time periods. Compared to regularized parametric portfolio policies, where firm characteristics directly determine individual portfolio weights, our mean-variance optimized portfolios obtain more than 50 percent higher out-of-sample Sharpe ratios and generate gains in alphas of more than 100 percent. The alphas of our estimated efficient portfolios are significant and amount to 32, 42 and 33 percent per annum with respect to the Fama-French six-factor model. The six-factor model is thus unable to explain the sizeable outperformance of our estimated efficient portfolios which could provide for estimates of the relevant stochastic discount factor that prices the considered assets.

We find that a linear ridge regression model applied to predict expected returns for mean-variance optimal portfolios is competitive with the nonlinear methods on our datasets. An ensemble of the linear model, gradient boosted regression trees and an artificial neural network improves out-of-sample Sharpe ratios by 8–10% compared to the linear specification. The realized performance of mean-variance optimized portfolios decreases over the more recent 2008 to 2020 subperiod for the US data. Sharpe ratios of the ensemble model then are 1.77, 2.05 and 2.30, i.e., gains in out-of-sample Sharpe ratio with respect to equally-weighted portfolios amount to 180, 250 and 390 percent, respectively. We apply the jackknife method on the estimated covariance matrix to deal with estimation risks from optimized portfolio weights aiming to target a prespecified risk limit for the portfolio's expected out-of-sample variance. Thereby out-of-sample volatilities of optimized portfolios are near the specified volatility target of 15% per annum and are comparable to a market index portfolio.

In our empirical application, monthly estimated mean-variance efficient portfolios produce an excessive amount of turnover averaging over 500 percent per month. Resulting transaction costs from monthly rebalancing might be able to wipe out the outperformance with respect to the market indices. We therefore analyze how measures of turnover reduction based on partial rebalancing affect the optimized portfolios' out-of-sample performance. A reduction of portfolio turnover to about 100 percent per month does not substantially impair realized performance. Turnover-reduced optimized portfolios exhibit Sharpe ratios of 2.24, 2.74 and 1.91 over the 1995–2020 and 2000–2020 out-of-sample periods, respectively. Our mean-variance optimized portfolios obtain gains in net Sharpe ratio of up to 280 percent compared to equally-weighted market portfolios after accounting for proportional round-trip transaction costs of 0.5%. We find significant economic value from estimated efficient portfolios as investors would be willing to switch to our portfolio strategy based on the ensemble prediction model unless they receive per annum 25 percent on the S&P 1500 and 19 percent on the S&P 500 and STOXX Europe 600 in addition to the respective market index return.

Future extensions of our methodology for estimating large mean-variance efficient portfolios could consider macroeconomic variables and further firm characteristics such as industry-specific characteristics or option-implied variables as in DeMiguel et al. (2013) for predicting expected returns. Applying more sophisticated models for the estimation of time-varying covariance matrices as in Engle et al. (2019) and De Nard et al. (2021) as well as specialized methods for using daily returns to forecast monthly covariance matrices are promising directions for further improvements of the mean-variance optimized portfolios. Because the stochastic discount factor is closely related to the efficient portfolio, the estimated efficient portfolios could be used as a basis to refine interpretable empirical factor models for asset pricing.

Appendix 3.A: Supplementary tables

Table 3.A.1: Description of 27 firm variables used in the empirical analysis.

Variable	Description	References
Size	Log of end of month market capitalization	Banz (1981), Fama and French (1992, 2008)
B/M	Book-to-market ratio as log of book value of common equity divided by market capitalization	Rosenberg et al. (1985), Fama and French (1992, 2008)
$\text{Ret}_{1,0}$	Stock return over the previous month	Jegadeesh and Titman (1993, 2001)
$\text{Ret}_{12,2}$	Stock return from month $t-12$ to $t-2$	
$\text{Ret}_{12,7}$	Stock return from month $t-12$ to $t-7$	
$\text{Ret}_{36,13}$	Stock return from month $t-36$ to $t-13$	
$\text{Ret}_{60,13}$	Stock return from month $t-60$ to $t-13$	
DY	Dividend payments per share over the previous 12 months divided by end of month share price	Litzenberger and Ramaswamy (1982), Lewellen (2015)
Vol	Log excess stock return volatility based on previous 60 months	Ang et al. (2006)
$\text{Vol}_{\text{daily}}$	Log excess stock return volatility based on daily returns within the previous month	
Beta	Slope coefficient of linear regression of excess market returns on excess stock returns based on previous 60 months	Fama and MacBeth (1973), Fama and French (1992, 2006)
$\text{Beta}_{\text{daily}}$	Slope coefficient of linear regression of excess market returns on excess stock returns based on daily returns within the previous month	
IVol	Log idiosyncratic volatility from Beta and excess market return volatility	Ang et al. (2006), Bali and Cakici (2008)
$\text{IVol}_{\text{daily}}$	Log idiosyncratic volatility from $\text{Beta}_{\text{daily}}$ and daily excess market return volatility	

Table 3.A.1 (continued).

Variable	Description	References
Turnover	Log of stock turnover by volume divided by number of shares outstanding	Datar et al. (1998), Lewellen (2015)
Investment	Relative change in total assets from $t-13$ to $t-1$	Cooper et al. (2008)
Debt/Price	Log of total debt divided by market capitalization	Bhandari (1988)
Sales/Price	Log of net sales or revenues divided by market capitalization	Barbee et al. (1996), Lewellen (2015)
SalesGrowth	Log change of net sales or revenues from $t-13$ to $t-1$	Lakonishok et al. (1994)
Earnings/Price	Net income after preferred dividends divided by market capitalization	Basu (1977, 1983), Fama and French (1992)
EarningsGrowth	Relative change in net income after preferred dividends from $t-13$ to $t-1$	Lakonishok et al. (1994)
GrossProfit	Net sales or revenues minus cost of goods sold divided by total assets	Novy-Marx (2013)
CF/Price	Funds from operations divided by market capitalization	Lakonishok et al. (1994)
CashRatio	Log of cash and equivalents divided by total assets	Palazzo (2012)
Accruals	Change in working capital from $t-13$ to $t-1$ divided by book value of common equity from $t-1$	Sloan (1996), Fama and French (2008)
Issues	Log change of split-adjusted number of shares outstanding from $t-36$ to $t-1$	Pontiff and Woodgate (2008), Lewellen (2015)
MAX	Maximum daily return within the previous month	Bali et al. (2011)

Table 3.A.2: Hyperparameters for machine learning methods derived from out-of-sample validation over six subperiods on three datasets.

S&P 500	1	2	3	4	5	6
Ridge Regression						
λ	0.000	0.002	0.016	0.082	0.068	0.000
Boosted Trees						
ν	0.005	0.005	0.005	0.005	0.005	0.005
Neural Network						
LR decay	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
λ	0.40	0.30	0.20	0.40	0.30	0.10
S&P 1500	**1**	**2**	**3**	**4**	**5**	**6**
Ridge Regression						
λ	0.000	0.034	0.040	0.144	0.098	0.000
Boosted Trees						
ν	0.007	0.005	0.004	0.004	0.005	0.006
Neural Network						
LR decay	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
λ	0.10	0.20	0.25	0.15	0.05	0.10
STOXX Europe 600	**1**	**2**	**3**	**4**	**5**	**6**
Ridge Regression						
λ	–	0.016	0.028	0.080	0.190	0.270
Boosted Trees						
ν	–	0.006	0.006	0.005	0.003	0.003
Neural Network						
LR decay	–	0.0005	0.0005	0.0005	0.0005	0.0005
λ	–	0.05	0.05	0.10	0.40	0.50

Note: The column numbers 1 to 5 correspond to the 5-year subperiods 01/1995–12/1999 until 01/2015–12/2019, 6 is the period 01/2020–12/2020. As we conduct an out-of-sample validation every five years (the month before the start of each subperiod) to determine hyperparameters for XGBoost and the neural network, the given hyperparameters are constant during the subperiods. The regularization parameter λ for ridge regression is determined every 12 months, so that we report the average λ value during each 5-year-period.
The random forest method is implemented via XGBoost and shares the following set of hyperparameters for each subperiod and dataset: number of parallel trees 500, maximum tree depth 15, minimum node size 300, subsampling 0.3 and column sampling 1.
Gradient boosted trees via XGBoost share the following set of hyperparameters: number of trees 5000, maximum tree depth 8, minimum node size 150, subsampling 0.1, column sampling 1 and regularization parameter $\lambda = 2000$.
Neural networks share the following set of hyperparameters: one layer with 128 units, dropout rate 0.25, Adam optimizer with learning rate (LR) 0.001 and exponential decay given in the table, batch size 32 and number of training epochs of 200.

Table 3.A.3: Statistics of standardized mean-variance optimized portfolio weights in quintile portfolios formed on expected return predictions from machine learning methods using firm characteristics for firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

	S&P 500					S&P 1500					STOXX Europe 600				
	LR	RF	XGB	NN	ENS	LR	RF	XGB	NN	ENS	LR	RF	XGB	NN	ENS
Quintile Portfolio 1															
Avg	−1.02	−1.00	−1.06	−1.05	−1.03	−1.03	−1.02	−1.11	−1.09	−1.05	−1.04	−1.02	−1.12	−1.07	−1.06
Std	0.75	0.75	0.76	0.80	0.76	0.74	0.76	0.73	0.79	0.75	0.77	0.75	0.74	0.77	0.76
Quintile Portfolio 2															
Avg	−0.38	−0.40	−0.39	−0.36	−0.38	−0.37	−0.36	−0.40	−0.37	−0.40	−0.39	−0.40	−0.40	−0.40	−0.41
Std	0.65	0.66	0.63	0.66	0.66	0.64	0.63	0.59	0.61	0.63	0.65	0.65	0.61	0.64	0.65
Quintile Portfolio 3															
Avg	0.02	0.01	0.02	0.06	0.03	0.02	0.01	0.02	0.02	0.01	0.01	0.00	0.04	0.02	0.02
Std	0.66	0.67	0.64	0.68	0.67	0.66	0.64	0.60	0.62	0.65	0.65	0.66	0.61	0.63	0.66
Quintile Portfolio 4															
Avg	0.43	0.43	0.43	0.44	0.44	0.42	0.36	0.42	0.42	0.43	0.40	0.41	0.44	0.41	0.42
Std	0.71	0.72	0.68	0.71	0.72	0.72	0.69	0.63	0.66	0.69	0.68	0.69	0.64	0.65	0.67
Quintile Portfolio 5															
Avg	0.95	0.97	1.00	0.91	0.94	0.96	1.01	1.07	1.01	1.01	1.02	1.02	1.04	1.03	1.02
Std	0.89	0.86	0.84	0.83	0.85	0.88	0.91	0.80	0.82	0.83	0.82	0.82	0.77	0.79	0.79
Correlation of $\widehat{\mu}_t$ with $\widehat{w}_t$															
Cor	0.69	0.69	0.73	0.68	0.69	0.70	0.72	0.77	0.74	0.73	0.73	0.73	0.78	0.75	0.74

Note: The table reports averages and standard deviations of monthly standardized mean-variance optimized portfolio weights from (3.8) for stocks sorted into quintile portfolios based on their predicted expected returns from a linear regression (LR), random forest (RF), gradient boosted regression trees (XGB), neural network (NN) and ensemble model (averaging the predictions from LR, XGB and NN). The last row reports time-series averages of monthly calculated correlations between predicted expected returns and mean-variance optimized portfolio weights. Expected returns and mean-variance optimal portfolio weights are estimated for months in the out-of-sample periods from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

Table 3.A.4: Portfolio performance statistics for mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices based on covariance matrices estimated from monthly return data.

Portfolio Statistics	EW	MINV	SMV	LR	RF	XGB	NN	ENS
S&P 500								
Mean	10.72	7.60	7.28	21.26	23.91	23.82	22.91	23.35
Std	16.44	11.21	16.33	13.36	14.07	13.37	12.72	13.02
Sharpe Ratio	0.65	0.68	0.45	1.59	1.70	1.78	1.80	1.79
S&P 1500								
Mean	11.87	9.06	5.66	24.24	30.24	33.42	29.34	29.90
Std	18.88	10.81	15.57	13.06	13.98	13.02	12.97	12.77
Sharpe Ratio	0.63	0.84	0.36	1.86	2.16	2.57	2.26	2.34
STOXX Europe 600								
Mean	7.02	6.84	3.49	18.37	17.76	19.46	19.88	19.87
Std	17.03	10.16	16.59	13.78	14.23	14.50	14.41	14.19
Sharpe Ratio	0.41	0.67	0.21	1.33	1.25	1.34	1.38	1.40

Note: The table reports annualized means, standard deviations (both in percent) and Sharpe ratios of realized portfolio returns for eight portfolio strategies. EW and MINV denote the equally-weighted portfolio and estimated minimum variance portfolio based on 12 years of monthly return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8) based on covariance matrices from MINV, where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

Table 3.A.5: Skewness and kurtosis coefficients of realized out-of-sample returns for mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
Skewness	−0.53	−1.04	−1.09	−0.12	−0.14	−0.18	−0.09	−0.08	−0.10	0.94
Kurtosis	5.51	5.99	7.16	3.00	2.87	3.19	3.24	2.98	2.94	7.37
S&P 1500										
Skewness	−0.41	−1.08	−1.36	0.06	−0.42	0.41	0.40	0.12	0.09	1.93
Kurtosis	5.90	5.63	7.55	4.30	3.68	6.45	6.95	3.63	4.00	13.71
STOXX Europe 600										
Skewness	−0.57	−0.59	−0.77	−0.05	−0.21	−0.19	−0.05	−0.22	−0.18	0.16
Kurtosis	5.13	3.58	3.75	3.17	3.04	3.10	3.35	3.75	3.44	8.45

Note: The table reports skewness and kurtosis coefficients of realized portfolio returns for ten portfolio strategies. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

Table 3.A.6: Correlation matrix of mean-variance optimized portfolio weights of firms in the S&P 500, S&P 1500 and STOXX Europe 600 indices.

Portfolio Statistics	MINVD	LR	RF	XGB	NN	ENS	PPP
S&P 500							
MINVD	1.00	0.21	0.22	0.15	0.18	0.20	0.11
LR	0.21	1.00	0.71	0.67	0.84	0.90	0.44
RF	0.22	0.71	1.00	0.81	0.77	0.84	0.36
XGB	0.15	0.67	0.81	1.00	0.76	0.90	0.34
NN	0.18	0.84	0.77	0.76	1.00	0.94	0.40
ENS	0.20	0.90	0.84	0.90	0.94	1.00	0.43
PPP	0.11	0.44	0.36	0.34	0.40	0.43	1.00
S&P 1500							
MINVD	1.00	0.30	0.31	0.23	0.25	0.27	0.14
LR	0.30	1.00	0.69	0.64	0.83	0.86	0.54
RF	0.31	0.69	1.00	0.79	0.72	0.81	0.43
XGB	0.23	0.64	0.79	1.00	0.76	0.90	0.41
NN	0.25	0.83	0.72	0.76	1.00	0.93	0.46
ENS	0.27	0.86	0.81	0.90	0.93	1.00	0.51
PPP	0.14	0.54	0.43	0.41	0.46	0.51	1.00
STOXX Europe 600							
MINVD	1.00	0.24	0.30	0.20	0.22	0.24	0.06
LR	0.24	1.00	0.71	0.70	0.81	0.91	0.58
RF	0.30	0.71	1.00	0.82	0.72	0.82	0.43
XGB	0.20	0.70	0.82	1.00	0.75	0.90	0.43
NN	0.22	0.81	0.72	0.75	1.00	0.94	0.49
ENS	0.24	0.91	0.82	0.90	0.94	1.00	0.54
PPP	0.06	0.58	0.43	0.43	0.49	0.54	1.00

Note: The table reports time-series averages of monthly calculated correlation matrices of mean-variance optimized portfolio weights from (3.8), where expected returns are estimated based on a linear regression (LR), random forest (RF), gradient boosted regression trees (XGB), neural network (NN) or ensemble model (ENS, averaging the predictions from LR, XGB and NN) using firm characteristics. MINVD denotes an estimated minimum variance portfolio based on 24 months of daily return data. PPP is the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. Portfolio weights are estimated for months in the out-of-sample periods from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

Table 3.A.7: Risk-free rate in excess of the value-weighted market index return so that investors are indifferent to mean-variance optimized portfolios of firms in the S&P 500, S&P 1500 and STOXX Europe 600 for different levels of tolerated monthly turnover η after accounting for proportional round-trip transaction costs of 0.5%.

	EW	MINV	MINVD	SMV	LR	RF	XGB	NN	ENS	PPP
S&P 500										
$\eta = 1/6$	0.49	0.34	−0.18	−6.34	15.91	14.82	13.37	14.71	15.59	12.75
$\eta = 1/3$	0.49	−1.34	−3.03	−10.00	19.31	18.00	16.51	18.32	19.35	12.47
$\eta = 1$	0.49	−2.02	−8.36	−14.56	15.70	16.17	15.09	16.23	17.39	10.69
S&P 1500										
$\eta = 1/6$	0.04	2.36	0.66	−2.37	17.22	16.79	16.95	17.64	18.27	18.56
$\eta = 1/3$	0.04	1.16	−2.18	−6.19	20.27	20.40	21.60	21.74	22.79	19.29
$\eta = 1$	0.04	0.63	−7.42	−14.04	20.85	20.91	24.36	23.69	25.46	16.17
STOXX Europe 600										
$\eta = 1/6$	2.78	7.02	9.25	−1.09	14.39	11.52	11.43	12.65	13.44	10.35
$\eta = 1/3$	2.78	4.61	6.39	−1.96	17.79	14.14	14.61	16.79	17.30	11.72
$\eta = 1$	2.78	3.31	1.50	−5.06	18.26	15.64	15.90	19.01	19.31	10.13

Note: The table reports the annualized rate (in percent) that is required in excess of the value-weighted market index return so that its net Sharpe ratio (after accounting for proportional round-trip transaction costs of 0.5%) equals the net Sharpe ratios of the nine realized portfolio strategy returns for different levels η of tolerated monthly turnover. This is equivalent to the gain in certainty equivalent from a portfolio strategy in the table (scaled such that its volatility is equal to the market volatility) compared to the market index. EW, MINV and MINVD denote the equally-weighted portfolio and estimated minimum variance portfolios based on 12 years of monthly or 24 months of daily return data. SMV, LR, RF, XGB, NN and ENS are the turnover-reduced mean-variance optimized portfolios as defined in (3.13) and (3.14) based on the mean-variance optimized portfolios from (3.8), where expected returns are estimated based on sample means, a linear regression, random forest, gradient boosted regression trees, neural network or ensemble model (averaging the predictions from LR, XGB and NN) using firm characteristics. PPP is the turnover-reduced version of the regularized parametric portfolio policy (from DeMiguel et al. (2020)) as defined in (3.11) with $\gamma = 20$. The out-of-sample periods are from 01/1995 to 12/2020 for the S&P 500 and 1500 indices and from 01/2000 to 12/2020 for the STOXX Europe 600 index.

Chapter 4

Large dynamic covariance matrices and portfolio selection with a heterogeneous autoregressive model

The content of this chapter is joint work with Igor Honig[1] and corresponds to the following working paper (i.e., the first revision version which is under review by the Journal of Banking & Finance after a revise and resubmit decision):
Honig, I. and Kircher, F. (2024). Large dynamic covariance matrices and portfolio selection with a heterogeneous autoregressive model. Working paper, University of Regensburg.

Abstract
We propose a novel framework for modeling large dynamic covariance matrices based on heterogeneous autoregressive volatility and correlation components. We address the problem of estimation risks by using nonlinear shrinkage methods, making our framework applicable in high dimensions. Our model is parsimonious, flexible, and can be estimated using standard least squares methods. We perform a comprehensive empirical out-of-sample analysis and find significant statistical and economic improvements over common benchmark models. For minimum variance portfolios with over a thousand stocks, the annualized portfolio standard deviation improves to 8.92% compared to 9.75% for a (R)DCC-type model.

Keywords: time-varying covariance matrix; high dimension; heterogeneous autoregressive model; minimum variance portfolio; Markowitz portfolio optimization.

JEL classification: C22, C51, C58, G11.

[1] Chair of Statistics and Risk Management, Faculty of Business, Economics and Management Information Systems, University of Regensburg, 93040 Regensburg, Germany. Email: igor.honig@ur.de.

4.1 Introduction

Multivariate volatility modeling and forecasting is central to a wide range of financial applications. An active area of research is devoted to portfolio optimization, dating back to the seminal work of Markowitz (1952). The Markowitz mean-variance model requires as input the vector of expected returns and the return covariance matrix. Empirical evidence suggests that the process of return (co)variances is time-varying and highly persistent. Existing econometric frameworks, such as multivariate GARCH models, are able to capture these phenomena for a variety of assets. However, problems arise in high dimensions, i.e., when the number of assets is large relative to the sample size. In particular, estimation risks emerge when model parameters can only be estimated imprecisely from historical data, leading to unreliable portfolio weights and suboptimal out-of-sample performance of the optimized portfolios. To be tractable in practice, methodological refinements are needed for the estimation of covariance matrices in high dimensions. At the same time, multivariate volatility models should be parsimoniously parameterized, yet capture the inherent properties of the covariance dynamics.

In this paper, we propose a new framework for modeling large dynamic covariance matrices that is flexible, simple to implement, and can be estimated using standard least squares methods. Our model enables direct forecasting of the covariance matrix at common portfolio rebalancing frequencies (e.g., monthly, quarterly, etc.), while fully exploiting the information contained in daily asset returns. Using nonlinear shrinkage methods, our approach is applicable in high-dimensional settings with more than a thousand assets, and effectively deals with the problem of parameter estimation risks. In an empirical out-of-sample analysis, we compare our model with several benchmarks. In particular, we consider the dynamic conditional correlation model with nonlinear shrinkage (DCC-NL) from Engle et al. (2019), and a modified version of the realized DCC model from Bollerslev et al. (2020), in which nonlinear shrinkage is applied to the long-run correlation matrix (RDCC-NL). We find that our new framework yields statistically and economically significant improvements over these models for the construction of minimum variance portfolios.

Our framework is motivated by the HAR model of Corsi (2009)—which is a popular choice for realized volatility modeling—and builds on insights from Bollerslev et al. (2018). Using daily asset returns, we construct multiple components based on realized volatilities that reflect economically motivated short- and long-term investor horizons. We model future monthly volatilities as a linear function of the lagged volatility components, and estimate their dynamics jointly via least squares. Thereby, we exploit the commonalities in the behavior of realized volatilities across financial assets as reported in Bollerslev et al. (2018). The component structure of our model enables a parsimonious approximation of the observed long memory properties of realized volatilities. Continuing this line of argument, we extend the HAR-type structure described above for modeling future monthly realized correlations. The combination of realized volatilities and correlations results in a versatile covariance matrix model, with lower computational cost than the competing (R)DCC-NL model. While HAR models have been extended to multivariate realized volatility modeling, they have not been employed in high dimensions. We thus contribute to the literature in two main aspects.

Our first contribution is a new framework which we label the "Multivariate Heterogeneous Exponential" (MHEx) model, extending the class of HExp models of Bollerslev et al. (2018) to the multivariate case. To increase the flexibility of our model, we rely on the decomposition of the asset covariance matrix into realized volatilities and correlations. This approach is inspired by the class of dynamic correlation models (as in Engle (2002)), but it has not yet been considered in HAR-type frameworks. We further impose smoothness on the volatility and correlation dynamics by applying an exponential weighting scheme in the construction of the lagged realized volatility and correlation components. Both enhancements provide more precise estimates of the covariance dynamics compared to the original HAR model. At the same time, our model retains the conceptual advantages of parsimony and ease of implementation.

Our second contribution is the inclusion of shrinkage estimators in the context of multivariate HAR models, making them applicable in large dimensions. Most applications of HAR models consider a small number of assets, say ten, where shrinkage is not essential. Shrinkage estimators for the covariance matrix exist in linear (Ledoit and Wolf, 2003, 2004a,b) as well as nonlinear (Ledoit and Wolf, 2012, 2017, 2020, 2022) forms, and have been devised to alleviate estimation risks by reducing the variance of the sample covariance matrix estimator at the trade-off of introducing bias. We apply the nonlinear shrinkage function from Ledoit and Wolf (2020) to the lagged realized correlation matrices, i.e., the components that determine future monthly realized correlations. As a result, our framework is able to handle high-dimensional settings with more than a thousand assets, which is of practical importance for the selection of individual stock portfolios.

In an empirical out-of-sample analysis, we demonstrate the effectiveness of our proposed framework, which compares favorably with common benchmark models. We consider data sets where the assets are individual stocks as well as the case where the assets are predefined portfolios, with asset universes ranging from ten to 1500. We construct minimum variance portfolios and find that the gains from using our dynamic MHEx model over a static covariance matrix model increase consistently with the number of assets. The proposed MHEx model outperforms the (R)DCC-NL model, achieving statistical significance at the 1% level for all data sets. For example, for 100 portfolios formed on size and book-to-market characteristics, the annualized standard deviation of minimum variance strategies improves from 9.49% (9.27%)

using the (R)DCC-NL to 8.62% using the MHEx model. We find that the dynamic models yield rather unstable and extreme portfolio weights for large stock portfolios. In particular, the optimized portfolios produce high monthly portfolio leverage and turnover, which commonly exceeds 200% and 100%, respectively. Hence, for a more practice-oriented evaluation, we consider leverage- and weight-constrained portfolios. With portfolio constraints, the relative outperformance of the MHEx over the (R)DCC-NL model grows consistently with the number of assets, and is statistically significant at the 1% level for all but the smallest data set. For example, when investing in the S&P 1500 constituents, the standard deviation of 130/30 minimum variance portfolios—where 30% of the initial capital is in short positions—improves from 10.43% (9.75%) using the (R)DCC-NL to 8.92% using the MHEx model. To take an investor's risk-return trade-off into account, we additionally form mean-variance optimal portfolios. Our optimized portfolios using the MHEx model perform well relative to those using the static or (R)DCC-NL covariance matrix models, and significantly improve upon the equally-weighted benchmark portfolio across all data sets, yielding 67 to 114% higher out-of-sample Sharpe ratios. In summary, we conclude that our MHEx model adds value to the estimation of large dynamic covariance matrices, which can be used seamlessly for minimum variance and mean-variance portfolio construction.

Our paper crosses two main strands of literature. First, we draw on work that is concerned with the dynamic modeling of realized covariance matrices. Beginning with the seminal article of Andersen et al. (2003), the benefits of using measures for the ex-post (co)variation of asset returns have been studied intensively, motivated by the increased availability of high-frequency intraday data. Contributions along these lines include Gourieroux et al. (2009), Bauer and Vorkink (2011), Chiriac and Voev (2011), and more recent works of Opschoor et al. (2018) and Gribisch and Hartkopf (2023). While these frameworks share our approach of modeling realized covariance dynamics, the empirical applications focus exclusively on forecasting daily covariances and are limited to cases with relatively few assets. This raises doubts about the applicability of the models when the investment universe extends to several hundred securities, as is often the case in practice. Our approach differs from the above papers in that we compute monthly realized covariance matrices using daily data instead of intraday data, as the latter may not always be available for arbitrary asset universes.

Second, our paper extends existing work on modeling dynamic covariance matrices in large dimensions. To alleviate the curse of dimensionality, different methods have been proposed. In the context of dynamic correlation models, an important contribution is the DCC-NL model of Engle et al. (2019), which uses nonlinear shrinkage to estimate the long-run correlation matrix. De Nard et al. (2022) achieve further improvements for the DCC-NL model by incorporating intraday data. Another approach is to relate the covariance dynamics to common observable or latent systemic factors, see Bollerslev et al. (2019), Gribisch et al. (2020), or De Nard et al. (2021) for recent works. Factor models have also been used by Alves et al. (2023), who combine a vectorized HAR model with the least absolute shrinkage and selection operator (LASSO) to improve minimum variance portfolio estimates. Another relevant model class are factor copulas, which have been used in the works of Creal and Tsay (2015), Opschoor et al. (2021), and Oh and Patton (2023), among others. In this paper, we approach the problem from a different angle by proposing a parsimonious and flexible multivariate HAR-type modeling framework in conjunction with nonlinear shrinkage. Our approach is feasible in large dimensions without restricting to a factor structure and offers a useful alternative to existing econometric frameworks.

The paper is organized as follows. Section 4.2 describes our modeling and estimation framework for large dynamic covariance matrices and monthly portfolio selection. Section 4.3 details our empirical approach and discusses the main results of the out-of-sample portfolio analyses. Section 4.4 concludes.

4.2 Modeling large dynamic covariance matrices

The covariance matrix of asset returns is a key ingredient to determine Markowitz (1952) mean-variance optimal portfolio weights. In practice, this requires the estimation of a covariance matrix model which is often based on asset returns sampled at the daily frequency. In the context of portfolio selection, however, this leads to a temporal mismatch because portfolio weights are often updated at lower frequencies, e.g., monthly or quarterly. One way to address this issue is a forecast aggregation scheme applied to daily covariance matrix estimates up to the desired frequency, as proposed by De Nard et al. (2021). In this paper, we pursue a different approach, building on the empirical results of Ghysels et al. (2019), which suggest that a direct forecasting approach is the preferred choice for the class of models presented in this section.

Since the true covariance matrix is unobservable, we rely on realized measures constructed from higher frequency asset returns. For instance, the realized covariance matrix as an ex-post estimator of the true covariance matrix is consistent for ever finer sampling intervals, see Barndorff-Nielsen and Shephard (2004). As the primary objective of our empirical analysis is monthly portfolio selection over a long time period, we construct monthly realized measures based on daily asset returns. We thereby draw on the literature that considers realized (co)variation measures at the monthly frequency, as for portfolio construction and asset pricing, see, e.g., Bollerslev et al. (2022), Bekaert et al. (2022), or Moreira and Muir (2017).[2]

[2] In prior works, such measures of monthly realized (co)variances, going back to Merton (1980), have been considered in French et al. (1987), Brailsford and Faff (1996), Jagannathan and Ma (2003), and Ang et al. (2009).

Throughout the paper, we denote by $t = 1, \dots, T$ the monthly time index, where T is the sample size of monthly observations of a given data set. The subscript $i = 1, \dots, N_t$ indexes assets in month t, with N_t indicating that the number of investable assets varies over time. Monthly log returns are represented by the vector $r_t = (r_{1,t}, \dots, r_{N_t,t})'$, and daily log returns by $r_{t,k} = (r_{1,t,k}, \dots, r_{N_t,t,k})'$ for day k in month t.

4.2.1 Heterogeneous autoregressive volatilities and correlations

We consider a decomposition of the true but latent $(N_t \times N_t)$ monthly covariance matrix of asset returns $\Sigma_t = D_t R_t D_t$, where D_t is the diagonal matrix of N_t return volatilities, and R_t is the correlation matrix of asset returns. To build our model, we proceed analogously for the ex-post covariance matrix estimator for Σ_t, referred to as RCOV_t in the following, i.e.,

$$\text{RCOV}_t = \text{Diag}\left(\text{RV}_t\right) \text{RCOR}_t \, \text{Diag}\left(\text{RV}_t\right), \tag{4.1}$$

where $\text{Diag}\left(\text{RV}_t\right)$ denotes the diagonal matrix of N_t monthly realized volatilities, and RCOR_t is the monthly realized correlation matrix.

In each month t, we have N_t investable assets with daily log returns $r_{t,k}$ collected over a total of K_t days. We compute the vector of monthly realized volatilities as the sum of daily squared log returns,

$$\text{RV}_t = \sqrt{\sum_{k=1}^{K_t} r_{t,k}^2}, \tag{4.2}$$

i.e., we do not subtract the square of average daily returns. Instead, as is common in the finance literature, we assume that daily returns $r_{t,k}$ have an expected value of approximately zero and are uncorrelated. For the realized volatility dynamics, we consider a linear process

$$\text{RV}_t = \sum_{m \in \mathcal{M}_{\text{RV}}} \phi_m \, \text{ExpRV}_{t-1}^m + \varepsilon_t, \tag{4.3}$$

where ϕ_m, $m \in \mathcal{M}_{\text{RV}}$, are scalar parameters, and ε_t is an error term. $\mathcal{M}_{\text{RV}}$ denotes the set of different centers of mass for the exponential weights in the variables ExpRV_t^m as specified below. The model in (4.3) is formulated in terms of $|\mathcal{M}_{\text{RV}}|$ exponentially-weighted realized volatility measures, which are constructed based on past squared daily returns,

$$\text{ExpRV}_t^m = \sqrt{\overline{K} \sum_{l=0}^{L-1} \sum_{k=1}^{K_{t-l}} w_{t,j}^m \, r_{t-l,k}^2}, \tag{4.4}$$

where $j = \sum_{r=0}^{l} K_{t-r} - k$ indexes exponentially decreasing weights (for $j \to \infty$) for return observations in the past L months, including the current month. The weights in (4.4) are defined as

$$w_{t,j}^{m} = \frac{(1 + 1/m)^{-j}}{\sum_{k=0}^{\mathcal{K}_t - 1} (1 + 1/m)^{-k}}, \tag{4.5}$$

for $j = 0, \ldots, \mathcal{K}_t - 1$, where $\mathcal{K}_t = \sum_{l=0}^{L-1} K_{t-l}$ is the total number of days over the previous L months. The parameter m determines how quickly the exponential weights decay to zero. For large $\mathcal{K}_t$, m is equivalent to the center of mass defined as $\sum_{j=0}^{\mathcal{K}_t - 1} w_{t,j}^{m}\, j$ for weights $w_{t,j}^{m}$. Thus, the parameter m corresponds to the weighted average lag of days used to calculate ExpRV_t^m, reflecting the component's average time horizon.

For our empirical analysis, we choose common time horizons for the parameter m of the ExpRV_t^m variables, i.e., we specify $\mathcal{M}_{\text{RV}} = \{1, 5, 20, 60, 120, 250, \infty\}$. The special case $m = \infty$ implies equal-weighting of the squared daily returns in (4.4), hence ExpRV_t^∞ is equivalent to the mean of squared daily returns over the past L months. We choose $L = 60$, i.e., five years, and use ExpRV_t^∞ as the long-run volatility component, similar to Bollerslev et al. (2018). In general, the choice of $\mathcal{M}_{\text{RV}}$ depends on the research question at hand. We opt for a rather large number of components because we did not have problems with overfitting in our empirical analysis. To match the monthly horizon of the realized volatility RV_t modeled in (4.3), we apply a common scaling factor of $\overline{K} = 21$ in the definition of the ExpRV_t^m variables in (4.4), following the convention that there are on average $\overline{K} = 21$ trading days per month.

Our modeling approach for RV_t is inspired by the HExp model for intraday realized volatilities as considered in Bollerslev et al. (2018). The model resembles the structure of the univariate HAR model of Corsi (2009), which relies on a small number of predetermined volatility components to conveniently restrict the infinite-dimensional space of coefficients of a general AR(∞) process for RV_t. Although the HAR model has emerged as a widely used framework in many empirical studies, it has been refined to better capture important properties of financial time series. Bollerslev et al. (2018) propose a modification targeting the construction of the realized volatility components. In the original HAR model, the components are sample averages of realized volatilities over various time horizons (e.g., a day, a week, a month etc.), which results in a step-like function for the lagged RV_t's. In contrast, using exponential weights as in (4.4) achieves smoothness in the lag function for past RV_t's, with more recent daily observations having a greater impact on future volatility. Because only the construction of the model components is affected, and not the model structure itself, this approach maintains the benefits of the original HAR model. That is, our model in (4.3) can capture potential long memory behavior of return volatilities. Moreover, the model is simple to estimate and highly flexible due to the specified set of predetermined realized volatility components. For the multivariate modeling of realized volatilities, Bollerslev et al. (2018) find that volatilities of different financial assets behave similarly over time, so that imposing common model parameters yields out-of-sample forecasting improvements. Since this could be confirmed in our empirical application, we specify the model as in (4.3) with common parameters $\phi_m, m \in \mathcal{M}_{\text{RV}}$, for all assets.

We extend the above framework to the modeling of the dynamic realized correlation matrix RCOR_t. Analogous to the computation of realized volatilities in (4.2), we obtain the realized covariance matrix for month t from daily returns as

$$\text{RCOV}_t = \sum_{k=1}^{K_t} r_{t,k}\, r'_{t,k}. \tag{4.6}$$

Based on (4.1), we compute the monthly realized correlation matrix

$$\text{RCOR}_t = \text{Diag}\left(\text{RV}_t\right)^{-1}\ \text{RCOV}_t\ \text{Diag}\left(\text{RV}_t\right)^{-1}, \tag{4.7}$$

using (4.2) and (4.6). Due to the symmetry of the correlation matrix, only the dynamics of its lower triangular elements must be specified. We assume a linear process that is driven by lagged realized exponentially-weighted correlation components, i.e.,

$$\psi\left(\text{RCOR}_t\right) = \sum_{m \in \mathcal{M}_{\text{RCOR}}} \gamma_m\, \psi\left(\text{ExpRCOR}^m_{t-1}\right) + \nu_t, \tag{4.8}$$

where γ_m, $m \in \mathcal{M}_{\text{RCOR}}$, are scalar parameters, and ν_t is an error term. $\mathcal{M}_{\text{RCOR}}$ is the set of centers of mass for the realized correlation components, and $\psi(\cdot)$ denotes the operator stacking all elements below the main diagonal of the correlation matrix into an $N_t(N_t-1)/2$-dimensional vector. The regressors in (4.8) are defined as

$$\text{ExpRCOR}^m_t = \text{Diag}\left(\text{ExpRV}^m_t\right)^{-1}\ \text{ExpRCOV}^m_t\ \ \text{Diag}\left(\text{ExpRV}^m_t\right)^{-1} \tag{4.9}$$

with ExpRV^m_t as in (4.4) and

$$\text{ExpRCOV}^m_t = \overline{K} \sum_{l=0}^{L-1} \sum_{k=1}^{K_{t-l}} w^m_{t,j}\, r_{t-l,k}\, r'_{t-l,k}, \tag{4.10}$$

where $w^m_{t,j}$ is given in (4.5).

For the correlation model, we specify $\mathcal{M}_{\text{RCOR}} = \{10, 20, 60, 120, 250, \infty\}$, leaving out smaller values for m because they lead to rather noisy behavior of the corresponding ExpRCOV^m_t

components with $m < 10$. Similar to the ExpRV_t^m variables in (4.4), we scale the ExpRCOV_t^m variables in (4.10) by $\overline{K} = 21$ to match the monthly horizon of RCOV_t defined in (4.6).

By construction, the model assumes the same dynamics for all entries in the realized correlation matrix. This reduces the number of parameters to be estimated and makes our model (4.8) less prone to overfitting, similar to the DCC model (discussed in Section 4.2.4). Moreover, this guarantees positive definiteness of the correlation matrix for suitable restrictions on γ_m. Naturally, the correlation model (4.8) shares the main features of the volatility model (4.3). The use of multiple exponentially-weighted realized correlation components provides a parsimonious, flexible, and smooth specification of the dependence structure of future realized correlations upon their own past. The regression coefficients γ_m determine the relative influence, or weights, of the short- and long-term realized correlation components used to approximate the apparent long memory behavior of stock return correlations. We refer to the framework presented above, in which realized volatilities and correlations are modeled based on the decomposition of the realized covariance matrix, as the *"Multivariate Heterogeneous Exponential"*, or MHEx, model. Its denotation is derived from the HExp volatility model in Bollerslev et al. (2018) and its exponential weighting in the definition of the realized volatility and correlation components.

We identify three key advantages of our proposed framework. First, because the focus is on portfolio optimization, we require forecasts of the covariance matrix for the next month. We obtain these forecasts directly using (4.3), (4.8), and (4.1), thereby avoiding the need of a forecast aggregation scheme based on recursively obtained daily covariance matrix estimates, as applied in De Nard et al. (2021). Second, we achieve high flexibility by specifying the dependence structure of future realized volatilities based on multiple exponentially-weighted components covering a wide range of short-term and long-term realized volatility estimates. The same applies to the dynamic modeling of realized correlations, including the use of shrinkage estimators discussed in the next section to counteract the curse of dimensionality. Third, especially with respect to the realized correlation model, the use of predetermined components provides the advantage of simple implementation and computationally fast estimation. This is because the corresponding regression coefficients γ_m in (4.8) can be obtained with a suitable least squares estimator.

Remark 4.2.1. Our modeling framework is specifically formulated from the perspective of an investor who updates portfolio weights at a monthly frequency based on daily data. In fact, given (4.3) and (4.8), our methodology can be characterized as a mixed-frequency approach, since monthly realized volatilities and correlations are determined by variables measured at a higher (daily) frequency. Consequently, while our framework can be easily adapted to model covariance matrices at lower frequencies (e.g., quarterly, semi-annually, etc.), it cannot be used in exactly the same way for daily portfolio choice, which may be another relevant practical application. For such an application, our modeling framework needs to be modified to accommodate high-frequency intraday data. In practice, such data may not always be available for larger asset universes and, in particular, over long time periods as studied in the present paper.

4.2.2 Incorporating shrinkage estimators

It is well-known that in high dimensions—or for comparatively large N/T—the sample estimator of the covariance matrix, or the correlation matrix, is inadequate. This is because the number of free parameters in the matrix grows with N^2, allowing for reliable estimates only when T is an order of magnitude larger than N. Estimation errors become more severe for smaller T, however, they can be mitigated by adding structure to the covariance matrix. The attempt to find an optimal balance between bias and variance of a covariance matrix estimator has led to shrinkage estimators. Linear shrinkage estimators were introduced in Ledoit and Wolf (2004b) and extended to nonlinear shrinkage estimators, as in Ledoit and Wolf (2017) in the context of portfolio selection. Against this background, our goal is to make the MHEx model from Section 4.2.1 applicable to large portfolios. To do so, we propose a method to perform shrinkage on the exponentially-weighted (realized) correlation matrices.

Shrinkage estimators of the sample covariance matrix are developed under the assumption that observations are equally-weighted. We establish the connection between exponential and equal weighting of the data through the notion of effective sample size. Let $E_1 = \sum_{j=0}^{\mathcal{K}-1} w_j X_j$ be an estimator based on exponential weighting with weights w_j, and $E_2 = \frac{1}{\mathcal{J}} \sum_{j=0}^{\mathcal{J}-1} X_j$ be an estimator based on equal weighting, where X_j is a stationary series of serially uncorrelated random variables, which could, e.g., be the product of standardized returns as for a correlation measure. Shrinkage is designed to reduce estimation risk by shrinking the variance of an estimator in exchange for increased bias. Therefore, analyzing the unconditional variances of E_1 and E_2, we get that $\mathbb{V}\text{ar}(E_1) = \sigma_X^2 \sum_{j=0}^{\mathcal{K}-1} w_j^2 = \sigma_X^2/\mathcal{J} = \mathbb{V}\text{ar}(E_2)$ when $\mathcal{J} = 1/\sum_{j=0}^{\mathcal{K}-1} w_j^2$, where $\sigma_X^2 = \mathbb{V}\text{ar}\left(X_j\right)$ and $\mathbb{V}\text{ar}$ denotes the unconditional variance. The term $1/\sum_{j=0}^{\mathcal{K}-1} w_j^2$ represents the effective sample size for weights w_j, which gives the sample size of an equally-weighted statistic with the same variance or estimation risk as the w_j-weighted statistic. For the exponential weights in (4.5) we have that

$$\frac{1}{\sum_{j=0}^{\mathcal{K}_t-1} \left(w_{t,j}^m\right)^2} \to 2m+1 \text{ for } \mathcal{K}_t \to \infty.$$

Therefore, to apply nonlinear shrinkage to ExpRCOR_t^m in (4.9), we adopt the nonlinear shrinkage

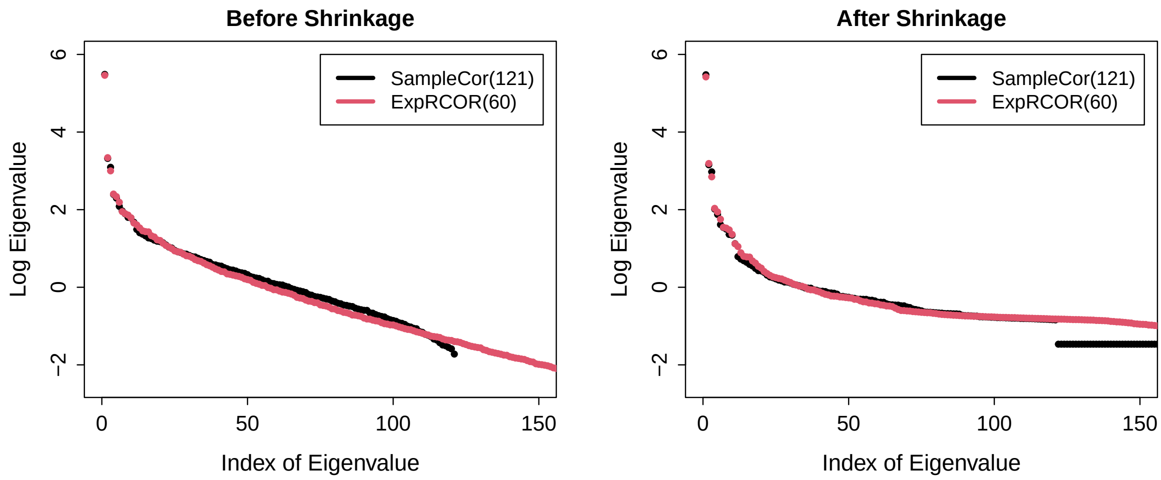

Figure 4.1: Left: log eigenvalues of the sample correlation matrix based on 121 observations and of ExpRCOR_t^m with $m = 60$. Right: log eigenvalues of the nonlinearly shrunk sample correlation matrix based on 121 observations and of ExpRCOR_t^m with $m = 60$ after applying the nonlinear shrinkage function.

function from the sample correlation matrix based on the past $2m + 1$ (equally-weighted) observations, which also imply a center of mass of m. The estimation risk and, as further shown below, the structure of the two correlation matrices based on the same data are comparable, indicating that the optimal shrinkage intensity is similar and the nonlinear shrinkage function can be seamlessly transferred to ExpRCOR_t^m.

In our framework, we use the analytical nonlinear shrinkage estimator of Ledoit and Wolf (2020), because it is computationally fast and performs well for both small and large samples. The main inputs for the Ledoit and Wolf (2020) nonlinear shrinkage estimator are the sample size relative to the matrix dimension and the vector of eigenvalues. The nonlinear shrinkage function transforms the sample eigenvalues based on a nonlinear function such that large eigenvalues tend to decrease while small eigenvalues increase, see Ledoit and Wolf (2017, 2020). Therefore, provided that the eigenvalue structure of both covariance matrices is reasonably similar, we can readily apply the eigenvalue transformation implied by the nonlinear shrinkage estimator of the sample correlation matrix based on $2m + 1$ observations to ExpRCOR_t^m.

Figure 4.1 illustrates this relationship. The left plot shows the 150 largest eigenvalues of the sample correlation matrix based on 121 observations for 492 firms in the S&P 500 index in 12/2022, with available data over the past five years, along with the ExpRCOR_t^m estimator with $m = 60$. The eigenvalue structure is noticeably similar and the largest eigenvalues are almost identical. Only the smaller eigenvalues of ExpRCOR_t^m decay much slower because we use a five-year estimation window, which means that observations older than 121 days are not discarded completely, as in the SampleCor(121) estimator, but receive successively less weight. The right plot shows the largest eigenvalues of both correlation matrices after nonlinear shrinkage. We apply the nonlinear estimator to standardized returns to shrink the sample correlation matrix based on $2m + 1 = 121$ observations. We linearly interpolate, as well as extrapolate for the largest eigenvalues, the estimated nonlinear shrinkage function, which maps the original eigenvalues of the sample correlation matrix to the eigenvalues of the nonlinearly shrunk sample correlation matrix. We then use this nonlinear shrinkage function to shrink ExpRCOR_t^m. The result, as shown in Figure 4.1, again reveals a similar eigenvalue structure compared to the shrunk sample correlation matrix and a slower decay for the small eigenvalues of the shrunk ExpRCOR_t^m matrix. We note that the eigenvectors of ExpRCOR_t^m remain unchanged, as is the case in the nonlinear shrinkage approach of Ledoit and Wolf (2017, 2020). We employ this shrinkage methodology for all ExpRCOR_t^m, as defined in (4.9), in our practical applications of the MHEx model.

4.2.3 Parameter estimation

In the following, we address the parameter estimation of the proposed MHEx model. For each month t, the volatility model (4.3) is estimated via least squares with the following constraints on the regression coefficients, i.e., the optimization problem for $\phi = (\phi_m)_{m \in \mathcal{M}_{\text{RV}}}$ is

$$\begin{aligned} \min_{\phi} \sum_{s=t-T_{\text{E}}}^{t-1} & \left(\text{RV}_s - \sum_{m \in \mathcal{M}_{\text{RV}}} \phi_m \, \text{ExpRV}_{s-1}^m \right)' \\ & \left(\text{RV}_s - \sum_{m \in \mathcal{M}_{\text{RV}}} \phi_m \, \text{ExpRV}_{s-1}^m \right) \\ \text{subject to} & \sum_{m \in \mathcal{M}_{\text{RV}}} \phi_m = 1, \quad \phi_m \geq 0 \; \forall m \in \mathcal{M}_{\text{RV}}, \end{aligned} \tag{4.11}$$

where T_{E} denotes the sample size used to estimate the MHEx model. While the ExpRV_t^m variables are all non-negative, we must also restrict the coefficients ϕ_m to be non-negative to ensure weakly positive estimates of volatilities. The sum-to-one constraint for ϕ_m, $m \in \mathcal{M}_{\text{RV}}$, facilitates an interpretation in terms of forecast averaging, where the ExpRV_t^m can be viewed as different individual forecasts of future volatility combined by weighted averaging.

We then estimate the correlation model (4.8) via least squares with the following constraints on the regression coefficients, i.e., the optimization problem for $\gamma = (\gamma_m)_{m \in \mathcal{M}_{\text{RCOR}}}$ is

$$\begin{aligned} \min_{\gamma} \; & \sum_{s=t-T_{\text{E}}}^{t-1} \left(\psi\left(\text{RCOR}_s\right) - \sum_{m \,\in\, \mathcal{M}_{\text{RCOR}}} \gamma_m \, \psi\left(\text{ExpRCOR}_{s-1}^m\right) \right)' \\ & \left(\psi\left(\text{RCOR}_s\right) - \sum_{m \,\in\, \mathcal{M}_{\text{RCOR}}} \gamma_m \, \psi\left(\text{ExpRCOR}_{s-1}^m\right) \right) \\ \text{subject to} \; & \sum_{m \,\in\, \mathcal{M}_{\text{RCOR}}} \gamma_m = 1, \quad \gamma_m \geq 0 \; \forall m \in \mathcal{M}_{\text{RCOR}}. \end{aligned} \tag{4.12}$$

We estimate the model with a non-negativity and sum-to-one constraint on the parameters γ_m.[3] In (4.12) and (4.13), ExpRCOR_t^m refers to the exponentially-weighted realized correlation matrix (4.9) after application of the nonlinear shrinkage methodology described in Section 4.2.2. The number of observations in $\psi\left(\text{ExpRCOR}_t^m\right)$ grows rapidly in high dimensions (i.e., for large N_t), resulting in a huge sample size for the least squares estimator (4.12). For numbers of assets in the hundreds we therefore replace the ψ operator with a $\widetilde{\psi}$ operator, stacking only the elements of the first off-diagonal into an $N_t - 1$ dimensional vector. The idea behind this is similar to the composite likelihood method, as used in Engle et al. (2019) and discussed in Section 4.2.4. In our application, the differences in the coefficient estimates for γ_m using $\widetilde{\psi}$ instead of ψ in (4.12) are marginal, but the runtime of the analysis improves considerably.

Given the estimated coefficients $\widehat{\phi}_m, m \in \mathcal{M}_{\text{RV}}$, and $\widehat{\gamma}_m, m \in \mathcal{M}_{\text{RCOR}}$, we predict monthly volatilities $\widehat{D}_t$, correlation matrices $\widehat{R}_t$, and covariance matrices $\widehat{\Sigma}_t$ as

$$\begin{aligned} \widehat{D}_t &\equiv \text{Diag}\left(\widehat{\text{RV}}_t\right) \;\text{ where }\; \widehat{\text{RV}}_t = \sum_{m \,\in\, \mathcal{M}_{\text{RV}}} \widehat{\phi}_m \, \text{ExpRV}_{t-1}^m, \\ \widehat{R}_t &\equiv \widehat{\text{RCOR}}_t = \sum_{m \,\in\, \mathcal{M}_{\text{RCOR}}} \widehat{\gamma}_m \, \text{ExpRCOR}_{t-1}^m \text{ and} \\ \widehat{\Sigma}_t &\equiv \widehat{D}_t \widehat{R}_t \widehat{D}_t. \end{aligned} \tag{4.13}$$

The estimated correlation matrix $\widehat{R}_t$ and the covariance matrix $\widehat{\Sigma}_t$ are guaranteed to be valid covariance matrices because, given the constraints on the coefficients $\widehat{\gamma}_m$, $\widehat{\text{RCOR}}_t$ is a convex combination of positive definite matrices.

[3] We have also considered the sparsity-encouraging LASSO, allowing the sum of coefficients γ_m to be less than one. The intuition here is to shrink the estimated correlations towards zero, which is a common approach to address estimation risks in large covariance matrices, similar to the linear shrinkage approach in Ledoit and Wolf (2004b). As this did not yield superior statistical and economic performance compared to the optimization problem formulated in (4.12), these results are not reported.

4.2.4 Comparison with alternative covariance matrix models

We now discuss two competing covariance matrix models considered in our empirical analysis and highlight their differences with our proposed framework.

The DCC-NL model

The DCC-NL model as proposed by Engle et al. (2019), and used in De Nard et al. (2021, 2022), is estimated in two steps via (quasi) maximum likelihood by assuming a multivariate normal distribution for the vector of daily asset returns. In the first step, the asset-specific daily conditional variances are estimated using a GARCH(1,1) model,

$$\sigma^2_{i,t,k} = \omega_i + \alpha_i\, r^2_{i,t,k-1} + \beta_i\, \sigma^2_{i,t,k-1}, \quad i = 1, \ldots, N_t,$$

with GARCH parameters $(\omega_i, \alpha_i, \beta_i)$, where we assume zero means for the daily log returns $r_{t,k}$.[4] In the second step, the dynamic conditional correlation matrix is estimated using correlation targeting as in Engle et al. (2019). That is, the DCC parameters (C, α, β) in

$$\begin{aligned} Q_{t,k} &= (1-\alpha-\beta)\, C + \alpha\, s_{t,k-1}s'_{t,k-1} + \beta\, Q_{t,k-1}, \\ R_{t,k} &= \text{Diag}\left(Q_{t,k}\right)^{-1/2}\, Q_{t,k}\, \text{Diag}\left(Q_{t,k}\right)^{-1/2} \end{aligned}$$

are estimated by first replacing the unconditional correlation matrix C by a nonlinear shrinkage estimate of the covariance matrix of the estimated standardized returns $\widehat{s}_{t,k} = r_{t,k}/\widehat{\sigma}_{t,k}$.

In large dimensions, Engle et al. (2019) employ the composite likelihood method (Pakel et al. (2021)), which circumvents the time-consuming inversion of the conditional correlation matrix $R_{t,k}$ by assembling the likelihood for the full model from the likelihoods of adjacent pairs of assets only. Note that we take a similar approach when estimating the MHEx correlation model as described in Section 4.2.3, where only the correlations of adjacent pairs are considered. When estimating the DCC-NL model, we use the nonlinear shrinkage estimator from Ledoit and Wolf (2020) for $\widehat{C}$, and the composite likelihood whenever $N_t > 100$.

To obtain monthly covariance matrix forecasts, we follow De Nard et al. (2021, 2022) and make multi-step forward predictions for each day k within the next month, assuming $\overline{K} = 21$ days within a month. The covariance matrix forecast for the whole month t is then derived by aggregating the daily forecasts as in $\widehat{\Sigma}_t = \sum_{k=1}^{\overline{K}} \widehat{\Sigma}_{t,k}$ with $\widehat{\Sigma}_{t,k} = \widehat{D}_{t,k}\, \widehat{R}_{t,k}\, \widehat{D}_{t,k}$. Here, $\widehat{D}_{t,k}$ is the diagonal matrix of multi-step variance forecasts, and $\widehat{R}_{t,k}$ the multi-step correlation matrix

[4] For the first day of each month, when $k = 1$, the right-hand side (in slight abuse of notation) refers to the last day of the previous month.

forecast for day k.

To highlight the increased flexibility of our approach compared to the DCC-NL model, consider the first day of month t. The GARCH(1,1) model allows the ARCH(∞) representation

$$\sigma^2_{i,t,1} = \frac{\omega_i}{1-\beta_i} + \alpha_i \sum_{l=1}^{\infty} \sum_{s=1}^{K_{t-l}} \beta_i^j \, r^2_{i,t-l,s}, \qquad i = 1, \ldots, N_t,$$

with $j = \sum_{r=1}^{l} K_{t-r} - s$, for $\omega_i, \alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i < 1$. This equation can be interpreted as the long-run variance plus an exponentially-weighted realized variance component whose smoothing parameter or decay rate β_i is asset-specific and estimated along with the other model parameters. This is in contrast to our MHEx volatility model from Section 4.2.1, which contains multiple exponentially-weighted components with different pre-specified decay rates. Provided that these pre-specified parameters cover the space reasonably well, we gain in flexibility, allowing us to model the higher impact of more recent returns as well as the apparent long memory properties of realized volatilities. The GARCH(1,1) model has to compromise in this regard with only a single smoothing parameter. To some extent, this can be mitigated by considering higher order or multi-component GARCH models. However, model selection and estimation then becomes increasingly cumbersome. The same logic applies to the DCC model for dynamic correlations compared to our MHEx correlation model from Section 4.2.1.

The RDCC-NL model

While the DCC-NL model is specified on daily returns, and monthly forecasts are derived by aggregating daily forecasts over the next month, a DCC-type model estimated directly on monthly returns could be preferable given our focus on monthly portfolio selection.[5] DCC-type models, which include high-frequency data to improve their forecasting ability, are often called Realized DCC (RDCC) models, as in Bollerslev et al. (2020) and Bauwens and Xu (2023). We define a RDCC-NL model analogously to Bollerslev et al. (2020), however, we specify the model on monthly returns, use daily data within the model (instead of intraday data), and apply a nonlinear shrinkage estimator for the correlation matrix. The RDCC-NL model yields direct monthly covariance matrix forecasts based on daily data, thus allowing a more direct comparison with our MHEx model from Section 4.2.1. The RDCC-NL model is defined by the following

[5] We thank anonymous referees for the suggestion to consider a (Realized) DCC-NL model applied to monthly returns.

equations:

$$\begin{aligned}
\sigma_t^2 &= (1-\alpha_v-\beta_v)\ V + \alpha_v\,\mathrm{RV}_{t-1}^2 + \beta_v\,\sigma_{t-1}^2, \\
Q_t &= (1-\alpha_q-\beta_q)\,C + \alpha_q\,\mathrm{RCOR}_{t-1} + \beta_q\,Q_{t-1}, \\
R_t &= \mathrm{Diag}\,(Q_t)^{-1/2}\ Q_t\ \mathrm{Diag}\,(Q_t)^{-1/2},
\end{aligned} \tag{4.14}$$

where RV_t and RCOR_t are defined as in Section 4.2.1, V denotes the vector of long-run variances of the considered assets, C is the long-run correlation matrix, and σ_t^2 and R_t denote the monthly conditional variances and conditional correlation matrix, i.e., the conditional variances and correlation matrix of (simple) asset returns for month t.

In the RDCC-NL model specification, the lagged monthly realized volatilities RV_{t-1} (based on daily returns within month $t-1$) replace the lagged monthly returns that would be used in a DCC model specified on monthly data. Similarly, the lagged monthly realized correlation matrix RCOR_{t-1} replaces the outer product of lagged standardized monthly returns in the definition of the correlation dynamics. As a result, the RDCC-NL model uses the information from within-month returns more efficiently than a DCC model applied to monthly data, while allowing direct predictions of monthly covariance matrices. With the RDCC-NL model specification in (4.14), the unconditional variances and correlation matrix of monthly returns are essentially given by V and C, assuming that the unconditional levels of RV_t^2 and RCOR_t are approximately equal to V and C. Because the specification of the model on monthly returns implies that fewer observations are available for model estimation compared to the DCC-NL model estimated on daily returns, we specify α_q and β_q as common parameters for all assets in the variance equation in (4.14).

We estimate the RDCC-NL model similarly to the DCC-NL model via a two-step (quasi) maximum likelihood approach with variance and correlation targeting for V and C. That is, we estimate V as the mean of squared daily returns over the estimation period (scaled by $\overline{K}=21$ to conform to a monthly horizon), and we estimate C as the correlation matrix of daily returns via the nonlinear shrinkage estimator of Ledoit and Wolf (2020), which is applied to daily returns scaled to unit variance. In the first step, with $V$and C fixed, the common parameters α_v and β_v of the variance equation are estimated by maximizing the mean log likelihood of demeaned monthly (simple) returns of all assets. In the second step, the parameters α_q and β_q of the correlation equation are estimated based on the standardized (via the estimated volatilities from the previous step) monthly (simple) returns.[6] In large dimensions, such as for single stock

[6] In our empirical application, we have tested different specifications of the RDCC-NL model, i.e., additionally including the lagged (standardized) monthly returns in the specifications in (4.14), explicitly subtracting the mean of RCOR_{t-1} in the intercept specification, such as in $\left(1-\beta_q\right)C-\alpha_q\overline{\mathrm{RCOR}}$, estimating V and C via daily or monthly returns, and applying nonlinear shrinkage to RCOR_{t-1}. The specification described in this section proved to be the most effective.

data sets with $N_t > 100$, we use the composite likelihood method for parameter estimation of the RDCC-NL model, as described above for the DCC-NL model.

4.2.5 Portfolio optimization

We perform minimum variance and mean-variance portfolio optimization in our empirical analysis. As minimum variance portfolios require only the covariance matrix as an input, they provide a natural testing ground for covariance matrix forecasts, where the resulting portfolio strategies can then be evaluated based on their realized out-of-sample variances. The weights of the population minimum variance portfolio are given by

$$w_{\text{MINV},t} = \frac{\Sigma_t^{-1} 1_{N_t}}{1'_{N_t} \Sigma_t^{-1} 1_{N_t}}, \tag{4.15}$$

where 1_{N_t} denotes an $N_t \times 1$ vector of ones. The population covariance matrix Σ_t must be replaced by an estimate $\widehat{\Sigma}_t$ to obtain feasible portfolio weights $\widehat{w}_{\text{MINV},t}$ in practice.

When implementing (4.15), $\widehat{w}_{\text{MINV},t}$ can take large positive and negative values. This is often undesirable because it leads to a high concentration in certain assets. Therefore, portfolio optimization is usually implemented with constraints on the resulting portfolio weights. The minimum variance portfolio optimization problem with a leverage constraint and bounds on portfolio weights can only be solved numerically and is formulated as

$$\begin{aligned} &\min_{w} w' \Sigma_t w && (4.16) \\ \text{subject to} \quad & w' 1_{N_t} = 1, \ \sum_{i=1}^{N_t} |w_i| \leq \ell \text{ and } w_i \in [w_{\min}, w_{\max}] \, , \\ & i = 1, \ldots, N_t, \end{aligned}$$

where ℓ is the maximum gross leverage allowed, and $w_{\min}$ and $w_{\max}$ are the lower and upper bounds on portfolio weights, respectively.

As minimum variance portfolios are agnostic about expected returns, i.e., they implicitly assume that all assets have identical expected returns, mean-variance optimal portfolios are of interest because they exploit the trade-off between risk and return. We consider the mean-variance portfolio optimization problem with a leverage constraint and bounds on portfolio weights,

$$\begin{aligned} &\min_{w} w'\mu_t - \frac{\gamma}{2} w'\Sigma_t w \\ \text{subject to} \quad & w'1_{N_t} = 1, \ \sum_{i=1}^{N_t} |w_i| \le \ell \text{ and } w_i \in [w_{\min}, w_{\max}], \\ & i = 1, \ldots, N_t, \end{aligned} \tag{4.17}$$

where γ is the coefficient of relative risk aversion. Without the leverage constraint and bounds on portfolio weights, (4.17) has the analytical solution

$$w_{\mathrm{MV},t} = w_{\mathrm{MINV},t} + \frac{1}{\gamma} \Sigma_t^{-1} \left(\mu_t - \mu_{\mathrm{MINV},t} 1_{N_t} \right), \tag{4.18}$$

where $\mu_{\mathrm{MINV},t} = w'_{\mathrm{MINV},t} \mu_t$ denotes the expected return of the population minimum variance portfolio. To implement (4.17) in practice, we require a model for the vector of expected returns μ_t. To this end, we follow Lewellen (2015) and formulate a linear predictive regression model for μ_t based on firm-individual return reversal, return momentum, size and value variables. The goal is not to build the most advanced mean-variance optimal portfolios as there are many more predictive variables and highly complex methods from the literature worth considering. Rather, we want to examine the performance of the predicted dynamic covariance matrices in the context of 'full' mean-variance optimized portfolios based on a useful model for predicting expected returns (see Lewellen (2015)). Inserting the predictions or estimates $\widehat{\mu}_t$, $\widehat{\Sigma}_t$, and $\widehat{w}_{\mathrm{MINV},t}$ into (4.18) yields estimated mean-variance optimal portfolio weights $\widehat{w}_{\mathrm{MV},t}$ that depend on the parameter γ, which we determine empirically as described in Section 4.3.2.

Remark 4.2.2. The MHEx and DCC-NL approaches provide predictions for covariance matrices of monthly log returns, because they rely on daily log returns summing to monthly log returns. However, for portfolio optimization as described in this section, Σ_t represents the covariance matrix of simple returns. In our practical application, the predicted covariance matrix of log returns provides a suitable approximation for the covariance matrix of simple returns, since the differences are negligible relative to forecasting errors and the typical differences between the covariances of different equities.[7] Thus, we use the predictions of the MHEx and DCC-NL approaches defined in the preceding sections as $\widehat{\Sigma}_t$.

[7] For example, assuming a multivariate normal distribution for monthly log returns with expected returns of 0.005 and standard deviations of 0.05 (i.e., 6% and 17% per annum) and a correlation of 0.5, the relative difference between the simple return covariance and the log return covariance is 1.4% (based on the properties of the lognormal distribution). As this affects all assets to some extent, it has virtually no effect on the estimated portfolio weights.

4.3 Empirical analysis

Our empirical analysis is designed as a (pseudo) out-of-sample test of the covariance matrix models described in Section 4.2. Because our application is portfolio optimization, the primary objective of the analysis is to compare the out-of-sample performance of estimated minimum variance portfolios. This is a natural testing ground for the accuracy of dynamic covariance matrix estimators as the portfolios are formed solely on the basis of these predicted covariance matrices. We perform backtesting exercises for five data sets with an increasing number of assets ranging from ten to over a thousand to evaluate the usefulness of the MHEx model from Section 4.2.1 for a wide range of potential applications.

4.3.1 Datasets

Our empirical analysis includes five data sets with corresponding time periods given in Table 4.1. The first three (FF) portfolio data sets are available on Kenneth R. French's website, and the last two stock market index-based (SP) data sets are from Refinitiv Datastream. For each data set, we obtain data on daily total returns and monthly size and book-to-market equity measures.

As shown in Table 4.1, we conduct the out-of-sample analysis for applications to predefined portfolios, which implies a smaller number of assets to be modeled, such as ten or 30 industry portfolios or 100 stock characteristics-based portfolios, and for applications to single stocks, where the number of individual assets can be very large, such as the constituents of the S&P 500 and S&P 1500 indices.

Since the FF100 data set contains a small number of missing values at the beginning of the sample, we replace each missing value with the average value of the neighboring portfolios. End-of-month constituent lists for the index-based data sets are available starting in 09/1989 for the S&P 500 and in 12/1994 for the S&P 1500. For the months prior to that, we use the first available constituent list. This only affects the estimation periods, where the first constituent list is always included, while the out-of-sample test period begins after the first list is available.

The book equity data for firms receive a time lag of 4 months (as in Lewellen (2015)), after which we assume that they are publicly available to calculate book-to-market ratios for predicting the expected returns as described in Section 4.2.5. Before estimating the expected return model for the SP data sets, we cross-sectionally winsorize the predictor as well as the predicted variables at the 3% level, i.e., observations below the 3%-quantile or above the 97%-quantile are set equal to the respective quantile.

We perform portfolio optimization using only stocks that are index constituents, i.e., the portfolio built for month t consists of the index constituents at the end of month $t-1$. This

Table 4.1: Data sets used in the empirical analysis, with full time period and start of the out-of-sample (OOS) test period.

#	Data set	Abbreviation	Time Period	Start OOS Period
1	10 US industry portfolios	FF10	01/1950–12/2022	01/1970
2	30 US industry portfolios	FF30	01/1950–12/2022	01/1970
3	100 US Size-BM portfolios	FF100	01/1950–12/2022	01/1970
4	S&P 500 index constituents	SP500	01/1980–12/2022	01/1995
5	S&P 1500 index constituents	SP1500	01/1980–12/2022	01/1995

approach avoids a look-ahead bias and excludes micro-cap and illiquid stocks from investment. We require the stocks to have a complete daily return history for the past 60 months in order to make predictions with the MHEx model and to estimate the (R)DCC model. Due to this restriction, the number of stocks in the portfolio is generally less than 500 and 1500 for the SP500 and SP1500 data sets, respectively. Throughout the out-of-sample period, our SP500 portfolio contains an average of 480 stocks, and our SP1500 portfolio contains an average of 1360 stocks. While initially there are less stocks available with the required data history, the number of stocks within the last 20 years of the out-of-sample period is close to the total number of index constituents.

4.3.2 Empirical approach for parameter estimation and portfolio strategies

The (pseudo) out-of-sample analysis is conducted as follows. Data sets are divided into an initial estimation period and an out-of-sample test period. For the FF data sets starting in 1950, the out-of-sample test period starts in 1970 and ends in 2022, for the SP data sets, the out-of-sample test period starts in 1995 and ends in 2022 (see Table 4.1). Given the 5 years required to calculate the variables for the MHEx model, initial estimation periods are 15 and 10 years, respectively. We extend the estimation window each month until the window reaches its final size of $T_{\mathrm{E}} = 360$ months, i.e., 30 years. We estimate the MHEx model from Section 4.2.1 before the start of each month of the out-of-sample test period as described in Section 4.2.3. We predict volatilities and correlations and compute the covariance matrix as given in (4.13). The predicted covariance matrix is used as an input for portfolio optimization. We determine minimum variance portfolios without constraints on portfolio weights as in (4.15). In practice, it may be desirable to limit the size of individual portfolio positions or to impose restrictions on short sales. Therefore, we also perform the analysis for constrained minimum variance portfolios.

To further explore the usefulness of the dynamic covariance matrix models for portfolio optimization, we construct mean-variance optimal portfolios according to (4.17) to exploit the risk-return trade-off. For expected returns, we estimate a linear model each month using the same 30-year estimation window as for the MHEx covariance matrix model. We run a pooled time-series cross-sectional linear regression with an intercept, where the predictor variables, i.e., return reversal (stock return over the previous month), return momentum (stock return over the previous 12 months excluding the last month), size (log market capitalization), and value (book-to-market equity ratio), are centered each month with respect to their cross-sectional averages. We choose the coefficient of relative risk aversion, γ, such that the out-of-sample standard deviation of estimated portfolio weights $\widehat{w}_{\mathrm{MV},t}$ in (4.18) is approximately equal to the out-of-sample standard deviation of the equally-weighted portfolio. To do this empirically, we run a 10-fold cross-validation where we re-estimate the expected return model as well as the static and MHEx covariance matrix models to evaluate the out-of-sample portfolio performance on the omitted fold for a grid of γ values. We repeat this cross-validation procedure every 12 months. Because re-estimating large-dimensional (R)DCC models is time-consuming, we use the γ obtained for the MHEx model for all dynamic models.

4.3.3 Benchmark models

In our empirical analysis, we compare the MHEx approach for the estimation of large dynamic covariance matrices with other state-of-the-art methods:

- **Static** is a static covariance matrix estimator using the nonlinear covariance matrix shrinkage method from Ledoit and Wolf (2020) and a fixed estimation window as in Ledoit and Wolf (2017). We scale the static covariance matrix estimated from five years of daily returns (as for the DCC-NL model) by $\overline{K} = 21$ to obtain a prediction for the next month.

- **DCC-NL** is the well-established DCC model combined with a nonlinear shrinkage estimator for the unconditional covariance matrix, as introduced in Engle et al. (2019). Following Engle et al. (2019) and De Nard et al. (2021, 2022), we use the past five years of daily data to estimate the DCC-NL covariance model. We obtain a prediction for the next month from the DCC-NL model by aggregating the DCC-NL covariance matrix forecasts over $\overline{K} = 21$ days, following De Nard et al. (2021, 2022).

- **RDCC-NL** is a Realized DCC model analogously to Bollerslev et al. (2020), but applied to monthly returns, using daily data within the model and applying the nonlinear shrinkage estimator of Ledoit and Wolf (2020) to the correlation matrix. Because the RDCC-NL model is estimated based on monthly data, we use an estimation window of 30 years as for the MHEx model, see Section 4.3.2. The inital estimation periods at the beginning of the out-of-sample periods are 20 and 15 years for the FF and SP data sets, respectively, which are extended over time until the estimation periods reach 30 years.[8] The RDCC-NL model yields direct predictions of conditional variances and correlation matrices for the next month based on (4.14), which we combine to a forecast of the monthly covariance matrix.

- **EW** is an equally-weighted portfolio of assets which does not require any information about the covariance matrix and is an empirically hard to beat benchmark, e.g., according to DeMiguel et al. (2009).

4.3.4 Performance evaluation

We evaluate the goodness-of-fit of the estimated models via mean squared errors (and simlarly via mean absolute errors) for predicted volatilities, correlations, and covariances:

$$\frac{1}{T - t_{\text{oos}} + 1} \sum_{t=t_{\text{oos}}}^{T} \left(\widehat{\text{RV}}_t - \text{RV}_t\right)' \left(\widehat{\text{RV}}_t - \text{RV}_t\right) \; / \; N_t,$$

$$\frac{1}{T - t_{\text{oos}} + 1} \sum_{t=t_{\text{oos}}}^{T} ||\widehat{\text{RCOR}}_t - \text{RCOR}_t||_F^2 \; / \; N_t^2,$$

$$\frac{1}{T - t_{\text{oos}} + 1} \sum_{t=t_{\text{oos}}}^{T} ||\widehat{\Sigma}_t - \text{RCOV}_t||_F^2 \; / \; N_t^2,$$

where t_{oos} indexes the first month of the out-of-sample period, $\widehat{\text{RV}}_t$, $\widehat{\text{RCOR}}_t$ and $\widehat{\Sigma}_t$ are the predictions of volatilities, correlation and covariance matrices, as in (4.13) for the MHEx model, and $|| \cdot ||_F$ denotes the Frobenius norm of a matrix.

For each month t in the out-of-sample period, we calculate monthly realized portfolio returns as $\widetilde{r}_t^{\text{PF}} = \widehat{w}_t' \, \widetilde{r}_t$, where $\widehat{w}_t$ are estimated optimal portfolio weights for month t and $\widetilde{r}_t$ denotes the vector of monthly simple returns of the N_t assets in the portfolio. We calculate daily realized portfolio returns as $\widetilde{r}_{t,k}^{\text{PF}} = \widehat{w}_{t,k}' \, \widetilde{r}_{t,k}$ for days $k = 1, \ldots, K_t$, where $\widehat{w}_{t,k} = \frac{\widehat{w}_t \circ \prod_{j=1}^{k-1} \left(1 + \widetilde{r}_{t,j}\right)}{\widehat{w}_t' \left(\prod_{j=1}^{k-1} \left(1 + \widetilde{r}_{t,j}\right)\right)}$ reflects that the portfolio weights $\widehat{w}_t$ from the beginning of the month change slightly as each day of the month passes. $\widetilde{r}_{t,k}$ denote daily simple returns of the N_t assets in the portfolio, and $\circ$ is the Hadamard product which also applies to the product operator $\prod$.

As portfolio performance statistics, we calculate annualized average portfolio returns, standard

[8] For the SP data sets, we have ensured that the single stocks in the portfolio have a complete five-year return history (see Section 4.3.1). To estimate the model for more than five years, we divide the estimation period into five-year subperiods, in which the RDCC-NL model is specified for the index constituents at the end of each subperiod.

deviations and Sharpe ratios from daily returns as

$$\frac{252}{\sum_{t=t_{\text{oos}}}^{T} K_t} \sum_{t=t_{\text{oos}}}^{T} \sum_{k=1}^{K_t} \widetilde{r}_{t,k}^{\text{PF}}, \quad \sqrt{\frac{252}{\sum_{t=t_{\text{oos}}}^{T} K_t} \sum_{t=t_{\text{oos}}}^{T} \sum_{k=1}^{K_t} \left(\widetilde{r}_{t,k}^{\text{PF}} - \underline{\widetilde{r}}^{\text{PF}}\right)^2},$$

$$\frac{\frac{252}{\sum_{t=t_{\text{oos}}}^{T} K_t} \sum_{t=t_{\text{oos}}}^{T} \sum_{k=1}^{K_t} \widetilde{r}_{t,k}^{\text{PF}} - \widetilde{r}_{t,k}^{\text{Rf}}}{\sqrt{\frac{252}{\sum_{t=t_{\text{oos}}}^{T} K_t} \sum_{t=t_{\text{oos}}}^{T} \sum_{k=1}^{K_t} \left(\widetilde{r}_{t,k}^{\text{PF}} - \widetilde{r}_{t,k}^{\text{Rf}} - \left(\underline{\widetilde{r}}^{\text{PF}} - \underline{\widetilde{r}}^{\text{Rf}}\right)\right)^2}},$$

where the underline represents average returns and $\widetilde{r}_{t,k}^{\text{Rf}}$ denotes the daily risk-free rate which we obtain from Kenneth R. French's website.

For the estimated minimum variance portfolios we test if the differences in out-of-sample variances are statistically significant using the HAC inference method from Ledoit and Wolf (2011). For the mean-variance portfolios we use the HAC inference method from Ledoit and Wolf (2008) to test if the differences in out-of-sample Sharpe ratios are statistically significant. The primary comparison is between our dynamic MHEx model and the dynamic DCC-NL and RDCC-NL models.

To analyze the properties and implementability of the portfolio strategies, we calculate the average gross portfolio leverage, the average proportion of negative portfolio weights, and the average monthly portfolio turnover as

$$\frac{1}{T - t_{\text{oos}} + 1} \sum_{t=t_{\text{oos}}}^{T} \sum_{i=1}^{N_t} |\widehat{w}_{i,t}|, \quad \frac{1}{T - t_{\text{oos}} + 1} \sum_{t=t_{\text{oos}}}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} 1_{\{\widehat{w}_{i,t} < 0\}}, \tag{4.19}$$

$$\frac{1}{T - t_{\text{oos}}} \sum_{t=t_{\text{oos}}+1}^{T} \sum_{i=1}^{N_t} |\widehat{w}_{i,t} - \widehat{w}_{i,t-1}^{+}|,$$

where $\widehat{w}_{t-1}^{+}$ denotes the vector of portfolio weights at the end of month $t-1$ that result from holding the portfolio given by weights $\widehat{w}_{t-1}$ through month $t-1$ (for the SP data sets we also account for firms leaving and entering the portfolio).

4.3.5 Empirical results

In this section, we present empirical backtesting results of optimized portfolio strategies. First, we consider the goodness-of-fit statistics of the estimated dynamic covariance matrix models, second, we discuss the performance statistics and properties of unrestricted and restricted minimum variance and mean-variance optimal portfolios.

Table 4.2: Out-of-sample goodness-of-fit statistics for predicted volatilities, correlations, and covariances from the static and dynamic DCC-NL, RDCC-NL, and MHEx models.

	RMSE			MAE		
	Volatilities	Correlations	Covariances	Volatilities	Correlations	Covariances
			FF10			
Static	2.860	0.203	53.468	1.883	0.138	18.760
DCC-NL	2.151	0.181	**47.385**	1.290	0.126	**13.423**
RDCC-NL	2.236	0.189	49.568	1.291	0.131	13.623
MHEx	**2.144**	**0.174**	48.140	**1.272**	**0.120**	13.777
			FF30			
Static	3.175	0.219	60.096	2.075	0.162	21.136
DCC-NL	2.390	0.203	**52.222**	1.439	0.153	**15.084**
RDCC-NL	2.499	0.208	55.030	1.464	0.157	15.580
MHEx	**2.381**	**0.192**	53.134	**1.420**	**0.143**	15.574
			FF100			
Static	3.088	0.192	60.661	1.967	0.143	21.241
DCC-NL	2.351	0.178	**51.522**	1.393	0.136	**14.873**
RDCC-NL	2.489	0.195	56.081	1.491	0.152	16.074
MHEx	**2.323**	**0.171**	52.676	**1.356**	**0.127**	15.351
			SP500			
Static	6.272	0.259	105.888	3.942	0.209	37.997
DCC-NL	5.125	0.249	96.348	2.927	0.202	29.474
RDCC-NL	5.161	0.253	95.939	3.078	0.204	31.526
MHEx	**4.850**	**0.245**	**93.313**	**2.745**	**0.197**	**29.099**
			SP1500			
Static	7.216	0.255	113.234	4.697	0.207	43.849
DCC-NL	6.261	0.244	103.861	3.669	0.198	36.301
RDCC-NL	6.178	0.249	103.998	3.741	0.200	37.897
MHEx	**5.867**	**0.241**	**101.802**	**3.359**	**0.194**	**35.516**

Note: The table reports root mean squared errors (RMSE) and mean absolute errors (MAE) for the predictions of monthly realized volatilities, correlations, and covariances of percentage returns over the out-of-sample periods of each data set given in Table 4.1. Bold numbers indicate the lowest RMSE and MAE among the compared models.

Goodness-of-fit comparison

Table 4.2 presents goodness-of-fit measures for the static covariance model as well as the dynamic DCC-NL, RDCC-NL, and MHEx models. We report RMSE and MAE for volatilities, correlations, and covariances, since the (R)DCC-NL and MHEx models are all estimated via two-step approaches. In terms of goodness-of-fit statistics, a distinction can be made between the portfolio (FF) and single stock (SP) data sets. The more pronounced unpredictable risk, i.e., noise in single stock returns is reflected in overall higher RMSE and MAE. For the FF data sets, the RMSE for volatility (or correlation) predictions is between 2.10 and 3.20 (or 0.17 and 0.22), while for the SP data sets, the RMSE lies between 4.80 and 7.30 (or 0.24 and 0.26). When comparing the different models, the results show fairly similar goodness-of-fit measures for the dynamic (R)DCC-NL and MHEx models, and clear improvements over the static covariance matrix model. For volatility (correlation) forecasts, the RMSE and MAE of the dynamic models are about 20 to 30% (5 to 15%) lower than those of the static model. The improvements of the MHEx model over the static model for correlation predictions are greater for the FF data sets (11 to 14% relative improvement) than for the SP data sets (5 to 6% relative improvement). For the covariance forecasts, where the realized covariances are presumably more driven by outliers (compared to the volatilities or correlations), we find about 10% lower RMSE and 15 to 30% lower MAE for the dynamic models. Comparing the dynamic models, the MHEx model has the lowest RMSE and MAE for volatility and correlation predictions for all five data sets; only for covariance forecasts on the FF data sets, the DCC-NL model has the lowest RMSE (MAE). In summary, the MHEx model ranks first in terms of goodness-of-fit measures, the DCC-NL model is a close second, and the RDCC-NL model is a clear third. Thus, the more flexible specification for modeling volatilities and correlations in the MHEx model proves to be beneficial. The anticipated advantage of direct monthly forecasts with the RDCC-NL model over the DCC-NL model could not outweigh its drawback of being estimated based on less frequent (monthly) returns.

Results for unrestricted minimum variance portfolios

Table 4.3 presents out-of-sample performance measures for estimated minimum variance portfolios without constraints on portfolio weights. We mainly focus on the performance results based on daily returns as they provide more accurate estimates of out-of-sample standard deviations, which we consider to be the primary performance measure for minimum variance portfolios. Across the five data sets, we find that all estimated minimum variance portfolios have significantly lower standard deviations than the equally-weighted portfolio, indicating that the use of covariance matrix estimates adds value within portfolio optimization. The annualized standard deviations of the equally-weighted portfolio are between 16.0 and 17.5% for the FF data sets and between 20.0 and 21.5% for the single stock SP data sets, while the standard deviations for the minimum variance portfolios are between 8.1 and 12.2%. The dynamic covariance matrix models, i.e., (R)DCC-NL and MHEx, lead to statistically significant improvements at the 1% level over the static model. In particular, MHEx significantly outperforms the static model with

Table 4.3: Out-of-sample performance measures of estimated minimum variance portfolios based on different covariance matrix models.

	Daily			Monthly		
	AV	SD	SR	AV	SD	SR
FF10						
EW	11.92	16.07	0.47	11.86	15.18	0.49
Static	12.02	12.20	0.63	12.16	12.98	0.60
DCC-NL	11.75	11.86	0.63	11.96	12.97	0.59
RDCC-NL	11.18	11.88	0.58	11.40	13.17	0.54
MHEx	11.54	**11.53*****	0.63	11.82	13.22	0.57
FF30						
EW	11.93	16.83	0.45	12.03	16.93	0.45
Static	10.66	11.45	0.55	10.87	12.57	0.52
DCC-NL	11.37	11.31	0.62	11.57	12.71	0.57
RDCC-NL	9.94	10.99	0.51	10.18	12.55	0.47
MHEx	10.27	**10.65*****	0.56	10.54	12.27	0.51
FF100						
EW	12.87	17.43	0.49	13.18	18.52	0.48
Static	17.83	9.67	1.40	18.51	13.21	1.07
DCC-NL	16.48	9.49	1.28	17.06	12.89	0.98
RDCC-NL	16.56	9.27	1.32	17.11	12.71	1.00
MHEx	16.10	**8.62*****	1.37	16.63	12.10	1.01
SP500						
EW	13.04	20.08	0.54	12.54	17.10	0.61
Static	8.88	11.30	0.60	8.92	11.41	0.60
DCC-NL	11.67	11.50	0.83	11.59	10.48	0.90
RDCC-NL	10.32	10.74	0.76	10.35	10.81	0.76
MHEx	11.33	**10.03*****	0.92	11.34	9.85	0.94
SP1500						
EW	15.20	21.44	0.61	13.62	18.83	0.61
Static	10.77	9.88	0.88	10.74	9.87	0.87
DCC-NL	11.94	9.22	1.07	11.96	9.15	1.07
RDCC-NL	11.05	8.78	1.02	11.23	9.15	0.99
MHEx	11.40	**8.10*****	1.15	11.58	8.62	1.10

Note: The table reports the annualized means (AV), standard deviations (SD), and Sharpe ratios (SR) of daily and monthly out-of-sample percentage returns of estimated minimum variance portfolios for each data set. Bold numbers indicate the lowest portfolio standard deviations based on daily returns. *, **, and *** indicate significant differences in standard deviations between the (R)DCC-NL and MHEx models at the 10%, 5%, and 1% levels.

absolute reductions in percentage standard deviations of 0.67, 0.80, 1.05, 1.27, and 1.78 across the five data sets. These improvements over the static model—e.g. from 9.67 to 8.62% for the FF100 data set—are economically significant and become larger as the number of assets and the diversification opportunities increase.

Comparing the results for the dynamic models, we find that the MHEx model outperforms the DCC-NL and RDCC-NL models for all data sets. The RDCC-NL outperforms or equals the DCC-NL model. For instance, the absolute reductions in annualized percentage standard deviations for the RDCC-NL versus the DCC-NL model are 0.22 for the FF100 and 0.45 for the SP1500 data set, respectively, with the former being statistically significant at the 1% level and the latter being statistically insignificant at the 10% level, mostly due to the smaller sample size. The MHEx model improves upon the RDCC-NL model with reductions in standard deviations of 0.36, 0.34, 0.66 for the FF data sets and 0.72, 0.68 for the SP data sets. That is, for unrestricted minimum variance portfolios, the gains over the RDCC-NL model are similar in magnitude for the two FF industry portfolio data sets, and are similar for the stock characteristics-based FF100 data set and the single stock SP data sets. All differences in out-of-sample standard deviations for the MHEx model are statistically significant at the 1% level and are economically meaningful. For example, for the SP1500 data set, the DCC-NL model yields an annualized standard deviation of 9.22%, the RDCC-NL model yields 8.78%, and the MHEx model yields 8.10%. The results show that the use of the MHEx model is beneficial for selecting from predefined portfolios as well as individual stocks.

Remark 4.3.1. Another way to enhance the compared models is by incorporating intraday data, which could be OHLC data as in the IDR-DCC-NL model in De Nard et al. (2022), or high-frequency intraday returns as in the RDCC model in Bollerslev et al. (2020) and the HExp model in Bollerslev et al. (2018). While for large individual stocks, OHLC data and intraday returns are available in recent subperiods of our data sets, we use readily available daily returns as the common basis for comparison in this paper. In the same way that the IDR-DCC-NL model yields improvements over the DCC-NL model—which may be as large as the improvements from using the MHEx model—we expect improvements for the MHEx model with OHLC data or intraday returns. Because our focus in this paper is on models with daily returns, we leave these extensions and analyses to future research.

For realized out-of-sample Sharpe ratios, we find that the estimated minimum variance portfolios based on the dynamic models outperform the equally-weighted portfolio for all data sets. A statistically significant outperformance over EW is achieved with the MHEx model for the FF100 characteristics portfolios at the 1% level and for the SP500 and SP1500 data sets at the 5% and 1% levels, respectively. Comparing the dynamic MHEx model with the static covariance

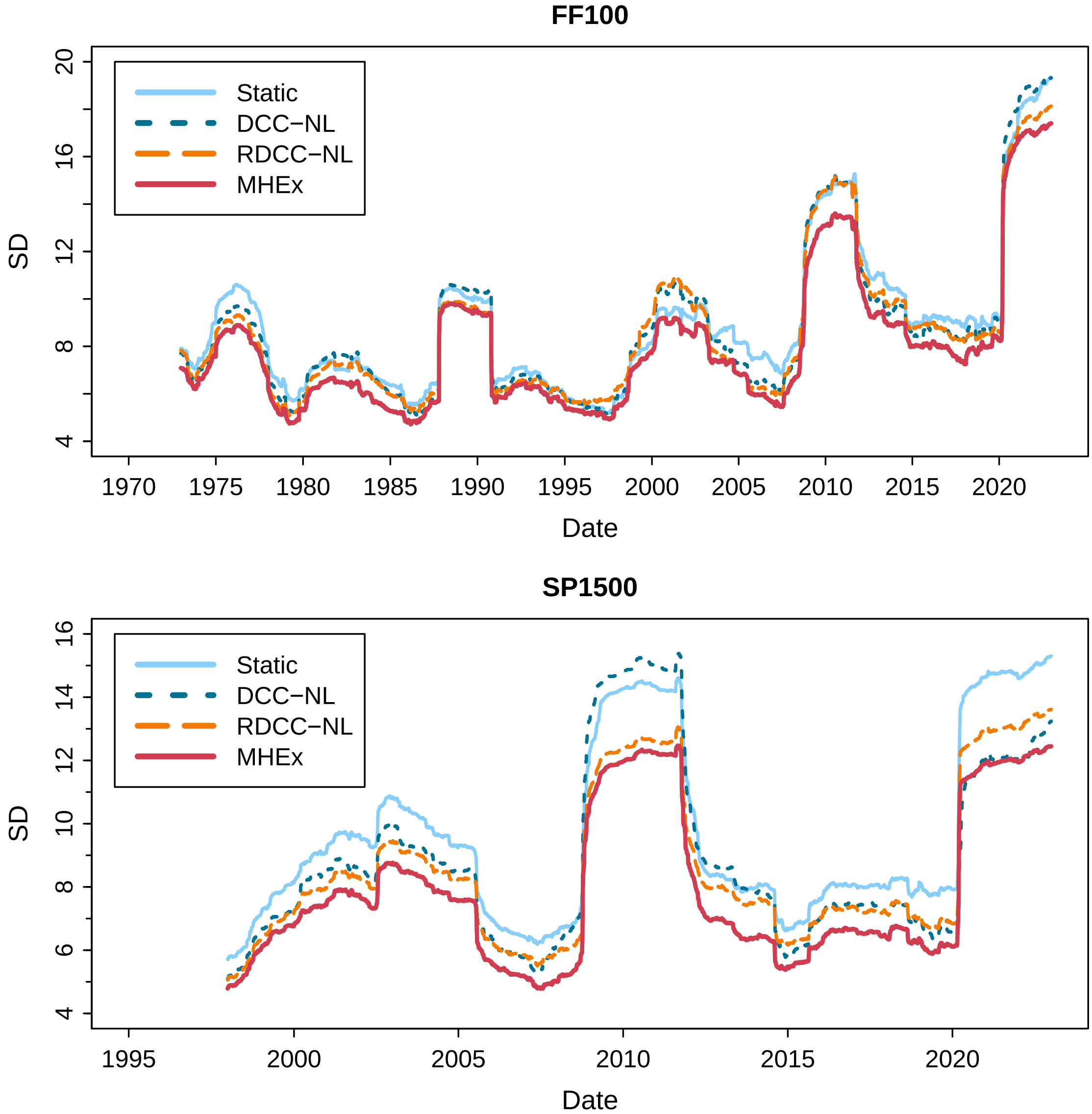

Figure 4.2: Annualized three-year rolling standard deviations of daily out-of-sample percentage portfolio returns of estimated minimum variance portfolios based on different covariance matrix models. The rolling window ends on the date indicated on the x-axis.

matrix model, the MHEx model outperforms the static model at the 1% significance level for the two single stock SP data sets. The out-of-sample Sharpe ratios are 0.92 (MHEx) and 0.60 (Static) for the SP500 data set and 1.15 (MHEx) and 0.88 (Static) for the SP1500 data set. For a more specific focus on Sharpe ratios, we discuss the results of mean-variance optimal portfolio selection with different dynamic covariance matrix models in Section 4.3.5.

To analyze the robustness of the empirical results in Table 4.3, we break down the aggregate performance measures into rolling subperiods of the considered out-of-sample periods. Figure 4.2 shows the annualized three-year rolling standard deviations for the estimated minimum variance portfolios using the static, (R)DCC-NL, or MHEx models for the investment universe of predefined portfolios, as for FF100, and individual stocks, as for SP1500. The MHEx model consistently outperforms the static and dynamic (RDCC-NL, DCC-NL) covariance matrix models over time. This holds true for various periods of low market volatility as well as for periods of high volatility, such as the early 2000s, the 2007/2008 financial crisis, or the 2020 stock market crash. Figure 4.2 shows that for practically all three-year subperiods between 1970 (1995) and 2020 for the FF100 (SP1500) data set, the MHEx model yields lower out-of-sample standard deviations than the static or (R)DCC-NL models. For the FF100 data set, there are several subperiods where the (R)DCC-NL model underperforms the static model, especially in periods of increased market volatility. In general, the RDCC-NL and DCC-NL models follow each other quite closely. For the SP1500 data set, the RDCC-NL model outperforms the DCC-NL model until 2015, while the DCC-NL model is superior after 2015. However, both are dominated by the MHEx model.

From this out-of-sample analysis, we conclude that the MHEx model is the preferred (high-dimensional) covariance matrix model, generally outperforming the RDCC-NL, DCC-NL, and static benchmark models for minimum variance portfolio optimization. The result is obtained for different investment universes consisting of predefined portfolios or large numbers of individual stocks for arbitrary subperiods of the respective out-of-sample periods. The next sections analyze the performance measures of mean-variance portfolios based on the compared covariance matrix models, the properties of the optimized portfolio weights, and the robustness of the portfolio performance statistics under reasonable weight constraints.

Results for unrestricted mean-variance portfolios

We construct mean-variance optimal portfolios to examine the economic value of the MHEx model for portfolios that seek to optimize a risk-return trade-off rather than focusing exclusively on risk minimization. Because the interaction between expected returns and covariance matrix is crucial in this context, the optimization problem is no longer clean for evaluating different covariance matrix models. Although we use the same expected return estimates in each case, we must interpret the results with some caution because optimal portfolio weights are quite sensitive to the expected return estimates, and a different expected return model may lead to different results.

Table 4.4 presents the out-of-sample performance statistics for the estimated unconstrained mean-variance optimal portfolios as given in (4.18). The parameter γ in (4.18) is determined via cross-validation so that the out-of-sample variances of estimated mean-variance optimal portfolios

Table 4.4: Out-of-sample performance measures of estimated mean-variance optimal portfolios based on different covariance matrix models.

	Daily			Monthly		
	AV	SD	SR	AV	SD	SR
			FF10			
EW	11.92	16.07	0.47	11.86	15.18	0.49
Static	17.99	16.36	0.84	18.35	17.86	0.79
DCC-NL	20.55	16.15	1.01	21.14	18.38	0.92
RDCC-NL	20.30	16.57	0.97	20.97	19.28	0.87
MHEx	19.56	**15.04**	**1.01**	20.10	17.26	0.92
			FF30			
EW	11.93	16.83	0.45	12.03	16.93	0.45
Static	17.70	16.26	0.82	17.93	17.63	0.77
DCC-NL	21.74	17.46	**1.00**	22.12	19.83	0.90
RDCC-NL	19.43	16.38	0.92	19.81	18.47	0.84
MHEx	19.59	**15.54**	0.98	20.00	17.82	0.88
			FF100			
EW	12.87	17.43	0.49	13.18	18.52	0.48
Static	21.11	17.15	0.98	23.21	22.11	0.85
DCC-NL	27.11	14.80	1.54	28.20	20.13	1.18
RDCC-NL	24.58	14.29	1.42	25.77	19.80	1.08
MHEx	26.58	**13.85**	**1.61**	27.83	20.43	1.15
			SP500			
EW	13.04	20.08	0.54	12.54	17.10	0.61
Static	17.30	**17.47**	0.87	17.53	18.01	0.86
DCC-NL	24.72	22.48	1.01	24.84	22.42	1.01
RDCC-NL	22.30	21.04	0.96	22.74	22.33	0.92
MHEx	23.10	19.28	**1.09**	23.47	20.31	1.05
			SP1500			
EW	15.20	21.44	0.61	13.62	18.83	0.61
Static	27.79	**19.06**	**1.35**	28.51	22.11	1.19
DCC-NL	30.51	26.00	1.09	31.05	26.98	1.07
RDCC-NL	27.88	22.66	1.14	28.82	25.24	1.06
MHEx	24.26	19.81	1.12	24.78	21.71	1.04

Note: The table reports the annualized means (AV), standard deviations (SD), and Sharpe ratios(SR) of daily and monthly out-of-sample percentage returns of estimated mean-variance portfolios for each data set. Bold numbers indicate the lowest portfolio standard deviations and highest Sharpe ratios based on daily returns.

for the MHEx and static models are approximately equal to those of the equally-weighted portfolio and can thus be compared more easily. We find that the so specified mean-variance portfolios have lower out-of-sample standard deviations than the EW portfolio, indicating that the cross-validation procedure yields conservative parameters in that the resulting portfolios are less risky than the target portfolio. Comparing MHEx and EW, the differences in standard deviations are statistically significant at the 1% level for the three FF data sets and at the 5% level for the SP1500 data set.

We evaluate the usefulness of the MHEx model for mean-variance portfolio optimization based on the Sharpe ratio, which is our primary performance measure of the risk-return trade-off. As shown in Table 4.4, the estimated mean-variance optimal portfolios using the MHEx model yield significantly higher out-of-sample Sharpe ratios than the equally-weighted market portfolio for all five data sets. The differences in Sharpe ratios are statistically significant at the 1% level for all data sets, except for SP1500 where the difference is significant at the 5% level. Investing in the FF10 industry portfolios, the FF100 size-book-to-market portfolios, or the SP1500 individual stocks yields annualized Sharpe ratios of 1.01, 1.61, 1.12 with mean-variance optimization using the MHEx model, compared to 0.47, 0.49, or 0.61 for the equally-weighted portfolio.

The dynamic MHEx model generates consistently higher out-of-sample Sharpe ratios for mean-variance optimized portfolios than the static model (except for the SP1500 data set). The differences in Sharpe ratios are statistically significant at the 1% level for the FF10 and FF100 data sets, and at the 5% level for the FF30 data set. However, since the static model benefits particularly from the estimated expected returns on the SP data sets, we no longer find the statistically significant outperformance for the MHEx model that was observed for the minimum variance portfolios.

Comparing the dynamic models, the differences in Sharpe ratios between MHEx and (R)DCC-NL are relatively small and not statistically significant. Based on the numbers, the MHEx model ranks first, the DCC-NL model second, and the RDCC-NL model third. That is, we find that overall the MHEx model is the favored covariance matrix model in this comparison and that it proves useful for potential mean-variance portfolio optimization.

Portfolio weight statistics for unrestricted portfolios

To analyze the empirical properties of the optimized portfolio weights, Table 4.5 presents statistics on the weights for the estimated unconstrained minimum variance and mean-variance optimal portfolios, namely the average maximum and minimum portfolio weights, proportion of negative portfolio weights, gross leverage, and portfolio turnover, as defined in (4.19). We find that the maximum and minimum portfolio weights are very similar across the different minimum

Table 4.5: Portfolio weight statistics for estimated minimum variance and mean-variance optimal portfolios based on different covariance matrix models.

	Minimum variance portfolios					Mean-variance portfolios				
	MAX	MIN	NEG	LEV	TO	MAX	MIN	NEG	LEV	TO
FF10										
EW	0.10	0.10	0.00	1.00	0.02	0.10	0.10	0.00	1.00	0.02
Static	0.77	−0.26	0.43	2.11	0.11	0.95	−0.57	0.44	3.35	2.15
DCC-NL	0.71	−0.26	0.42	2.08	1.13	1.04	−0.70	0.44	3.89	3.35
RDCC-NL	0.74	−0.20	0.44	1.82	0.63	0.99	−0.66	0.44	3.71	3.18
MHEx	0.76	−0.25	0.44	2.09	1.07	0.98	−0.61	0.44	3.63	2.93
FF30										
EW	0.03	0.03	0.00	1.00	0.03	0.03	0.03	0.00	1.00	0.03
Static	0.53	−0.19	0.46	2.84	0.21	0.59	−0.33	0.46	4.18	2.93
DCC-NL	0.50	−0.17	0.47	2.71	1.71	0.65	−0.40	0.47	4.95	4.51
RDCC-NL	0.58	−0.15	0.47	2.52	0.91	0.68	−0.37	0.47	4.57	3.95
MHEx	0.51	−0.15	0.47	2.66	1.56	0.62	−0.33	0.47	4.45	3.78
FF100										
EW	0.01	0.01	0.00	1.00	0.02	0.01	0.01	0.00	1.00	0.02
Static	0.32	−0.15	0.48	5.01	0.50	0.60	−0.27	0.50	8.22	4.71
DCC-NL	0.29	−0.11	0.50	4.58	3.26	0.78	−0.30	0.51	9.00	6.67
RDCC-NL	0.30	−0.12	0.50	4.63	2.34	0.71	−0.23	0.51	7.54	5.21
MHEx	0.29	−0.10	0.51	4.46	2.70	0.77	−0.26	0.52	8.76	5.57
SP500										
EW	0.00	0.00	0.00	1.00	0.07	0.00	0.00	0.00	1.00	0.07
Static	0.05	−0.04	0.44	4.69	1.53	0.15	−0.13	0.47	8.12	6.93
DCC-NL	0.13	−0.04	0.49	4.20	3.41	0.36	−0.20	0.50	12.86	10.28
RDCC-NL	0.09	−0.04	0.47	4.64	2.85	0.30	−0.22	0.49	13.36	10.83
MHEx	0.09	−0.03	0.48	3.74	2.62	0.21	−0.11	0.50	10.91	8.63
SP1500										
EW	0.00	0.00	0.00	1.00	0.09	0.00	0.00	0.00	1.00	0.09
Static	0.02	−0.01	0.42	4.58	2.70	0.04	−0.03	0.46	8.68	9.06
DCC-NL	0.10	−0.01	0.48	3.99	3.78	0.23	−0.08	0.51	16.32	15.62
RDCC-NL	0.05	−0.02	0.47	4.48	3.35	0.15	−0.07	0.50	16.36	16.33
MHEx	0.06	−0.01	0.48	3.81	3.03	0.14	−0.06	0.50	14.14	13.00

Note: The table reports the average maximum (MAX) and minimum (MIN) weights, proportion of negative weights (NEG), gross leverage (LEV), and turnover (TO) of estimated minimum variance and mean-variance portfolios for each data set.

variance and mean-variance portfolio strategies, generally decreasing in absolute value as the number of assets increases. Focusing on minimum variance portfolios, some portfolio positions are quite large, especially for the smaller portfolios, e.g., the largest long position is over 70% on average for the FF10 data set and over 9% on average for the SP500 data set. The largest short positions are about three times smaller than the largest long positions, but the proportion of negative portfolio weights is consistently between 42 and 50%, meaning that about half of the positions in the portfolios are short positions. Gross leverage is considerably higher for the optimized portfolios based on the FF100 stock characteristic portfolios and the SP single stock data sets than for those based on the FF10 and FF30 portfolios. For the former, the gross portfolio leverage ranges from 3.7 to 5.0, while for the latter it ranges from 1.8 to 2.8. The minimum variance portfolios based on the static model have slightly higher leverage than those based on the dynamic models. Using the MHEx model, the average gross leverage is lower than with the (R)DCC-NL model for the FF100 and SP data sets.

As expected, the equally-weighted portfolio has the lowest portfolio turnover in this comparison, as its portfolio positions are roughly the same from period to period. The minimum variance portfolios based on the static model display significantly lower turnover than those based on the dynamic covariance matrix models. The average monthly portfolio turnover generally increases with the number of assets, from 0.11 to 2.70 for the static model and from 1.07 to 3.03 for the MHEx model across the five data sets. Among the dynamic models, the RDCC-NL model yields the lowest turnover (it ranges from 0.63 to 3.35 across the five data sets) for the portfolio-based FF data sets, while the MHEx model yields the lowest turnover for the single stock SP data sets. The DCC-NL model has the highest turnover for all data sets.

Comparing the weight statistics of the estimated mean-variance optimal portfolios with those of the minimum variance portfolios, the mean-variance portfolios take on more leverage, have more extreme portfolio weights and higher portfolio turnover. The average proportion of negative portfolio weights is about the same at approximately 50%. For all data sets, the average gross leverage and monthly portfolio turnover are above 3.0 and 2.0, even for the smaller portfolios and with the static covariance matrix model. For the SP1500 data set, the gross leverage and portfolio turnover are very high, i.e., 8.7 and 9.1 on average with the static model, and considerably higher with the dynamic models. Among the dynamic models, the MHEx model produces the lowest leverage and turnover statistics for all data sets except FF100. For the FF10 portfolios, the average maximum portfolio weight is above 90% and the minimum weight below −60%. For the SP500 data set, the maximum weight is above 20% and the minimum weight below −10% on average. These figures illustrate the large long and short positions taken by the unconstrained minimum variance and mean-variance optimal portfolios, and the practical relevance of weight constraints to avoid such extreme portfolio positions.

Results for restricted portfolios

In terms of accuracy of dynamic covariance matrix estimation, the MHEx model compares favorably with the (R)DCC-NL model. This conclusion is based on goodness-of-fit statistics and the out-of-sample performance of estimated unrestricted minimum variance portfolios. The unrestricted portfolios have a large number of short positions, large absolute portfolio weights, high gross leverage, and high portfolio turnover, especially with the dynamic models. These characteristics render portfolio strategies relatively unattractive for practical implementation. To re-examine the application of the models in a more practical setting, we introduce constraints into the portfolio optimization problem, as formulated in (4.16) and (4.17), and compare the out-of-sample performance of the resulting estimated minimum variance and mean-variance optimal portfolios.

We limit the gross leverage to a maximum value of 1.6, which represents a 130/30 long-short equity strategy, i.e., the portfolio then holds no more than 30% of its initial capital in short positions. This type of strategy is quite popular among the long-short strategies of investment funds, as discussed in Lo and Patel (2008). We further impose restrictions on the size of individual portfolio positions, which is common among investment fonds. In our empirical application, portfolio positions are limited to 30% in absolute value for the FF10 and FF30 industry portfolios, 10% for the FF100 stock characteristic portfolios, and 5% for the single stock SP data sets.[9]

Table 4.6 presents the out-of-sample performance statistics for the constrained optimized portfolios. The standard deviations of the estimated minimum variance portfolios collectively rise as a result of the restrictions on portfolio weights. Considering the MHEx model, the smallest increase is from 10.65 to 11.36, i.e., 0.71 in terms of annualized percentage standard deviation, for the FF30 data set, while the largest increase is from 8.62 to 11.10, or 2.48, for the FF100 data set. These figures imply that the restrictions on portfolio leverage and the size of individual portfolio positions have an economically relevant impact.

The estimated minimum variance portfolios still have significantly lower out-of-sample standard deviations than the equally-weighted portfolio. The improvements in portfolio standard deviations with the MHEx model over the static model reduce slightly for the FF10 and FF30 industry portfolios, but remain robust for the FF100 and SP data sets when we compare the restricted with the unrestricted portfolios from Table 4.3. The improvements over the static model range from 0.36 to 2.14 and are statistically significant at the 1% level for the five data sets.

[9] We have also considered limiting gross portfolio leverage to 2.0 or 1.0, i.e., 150/50 long-short equity portfolios and long-only portfolios, with the same bounds on portfolio weights. The conclusions from these results are similar to those discussed in this section.

Table 4.6: Out-of-sample performance measures of estimated minimum variance and mean-variance optimal portfolios based on different covariance matrix models with a leverage constraint of 1.6 and data set-specific weight constraints.

	Minimum variance portfolios			Mean-variance portfolios		
	AV	SD	SR	AV	SD	SR
			FF10			
EW	11.92	16.07	0.47	11.92	16.07	0.47
Static	12.92	13.04	0.66	15.64	13.82	0.82
DCC-NL	12.08	12.83	0.61	15.62	13.81	0.82
RDCC-NL	11.85	12.77	0.59	15.86	13.87	**0.83**
MHEx	11.52	**12.68****	0.57	15.26	**13.62**	0.80
			FF30			
EW	11.93	16.83	0.45	11.93	16.83	0.45
Static	11.78	11.91	0.63	14.07	13.09	0.75
DCC-NL	11.79	11.75	0.64	15.15	13.43	**0.81**
RDCC-NL	10.54	11.57	0.54	15.12	13.29	**0.81**
MHEx	10.90	**11.36*****	0.58	14.44	**12.92**	0.78
			FF100			
EW	12.87	17.43	0.49	12.87	17.43	0.49
Static	16.41	12.31	0.98	16.47	12.88	0.94
DCC-NL	15.34	11.59	0.95	16.86	12.31	1.02
RDCC-NL	15.58	11.51	0.98	16.61	12.36	0.99
MHEx	15.23	**11.10*****	0.98	16.61	**11.74**	**1.05**
			SP500			
EW	13.04	20.08	0.54	13.04	20.08	0.54
Static	10.17	11.99	0.67	11.94	13.38	0.73
DCC-NL	10.45	11.88	0.70	13.68	13.57	0.85
RDCC-NL	9.99	11.32	0.70	12.62	12.75	0.82
MHEx	10.82	**10.87*****	0.80	13.27	**12.34**	**0.90**
			SP1500			
EW	15.20	21.44	0.61	15.20	21.44	0.61
Static	10.21	11.06	0.73	14.40	12.64	0.97
DCC-NL	11.40	10.43	0.89	15.83	12.99	1.06
RDCC-NL	10.16	9.75	0.82	14.21	11.62	1.04
MHEx	10.30	**8.92*****	0.92	13.46	**10.64**	**1.07**

Note: The table reports the annualized means (AV), standard deviations (SD), and Sharpe ratios (SR) of daily out-of-sample percentage returns of estimated minimum variance and mean-variance optimal portfolios for each data set. Bold numbers indicate the lowest portfolio standard deviations, or highest Sharpe ratios for the mean-variance portfolios, based on daily returns. *, **, and *** indicate significant differences in standard deviations between the (R)DCC-NL and MHEx models at the 10%, 5%, and 1% levels.

Comparing the dynamic models for restricted portfolios, the MHEx model outperforms the DCC-NL and RDCC-NL benchmark models for all five data sets. The improvements in terms of standard deviations are monotonically increasing with the number of assets, i.e., they are highest for the single stock data sets. For the FF30 and FF100 data sets, the MHEx model improves upon the RDCC-NL (DCC-NL) model by 0.21 (0.39) and 0.41 (0.49), while for the SP500 and SP1500 data sets, the MHEx model yields improvements of 0.45 (1.01) and 0.83 (1.51) in terms of annualized percentage standard deviations. Compared to the unrestricted portfolios, these differences in out-of-sample standard deviations between the (R)DCC-NL model and the MHEx model have decreased slightly (in the range of 0.12 to 0.27 with the RDCC-NL model), except for the SP1500 data set. Nonetheless, the differences in standard deviations are statistically significant at the 1% level and economically meaningful, especially for the single stock portfolios. For the SP1500 data set, the annualized out-of-sample standard deviations are 11.06% for the static model, 10.43% for DCC-NL, 9.75% for RDCC-NL, and 8.92% for MHEx. The results show that the MHEx model is the preferred high-dimensional covariance matrix model for minimum variance portfolios under portfolio constraints.

For the mean-variance optimized portfolios, the standard deviations in Table 4.6 decrease, indicating that out-of-sample standard deviations can be reduced by decreasing portfolio leverage and imposing weight constraints on the mean-variance portfolios. The weight constraints can have the additional effect of reducing the impact of estimation errors in $\widehat{\mu}$ and $\widehat{\Sigma}$ on the optimized portfolios. For this comparison, we do not recalibrate the γ parameter for (4.17) and use the same parameter as for the unrestricted mean-variance portfolios in Section 4.3.5.

Among the considered covariance matrix models, the MHEx model yields the lowest out-of-sample standard deviations, which are considerably lower than those of the equally-weighted portfolio. As shown in Table 4.6, the constrained mean-variance optimized portfolios consistently outperform the equally-weighted portfolio in terms of out-of-sample Sharpe ratios. The differences in Sharpe ratios between MHEx and EW are statistically significant at the 1% level for all data sets, except for SP500 where the difference is significant at the 5% level. For the FF10 industry portfolios, the FF100 size-book-to-market portfolios, or the SP1500 individual stocks, the MHEx approach yields Sharpe ratios of 0.80, 1.05, 1.07 versus 0.47, 0.49, 0.61 for equally-weighted portfolios, representing improvements of 70%, 114%, and 75%, respectively.

The Sharpe ratios for the constrained portfolios are generally lower than those for the unconstrained portfolios, i.e., by 17–35% for the MHEx approach, except for the SP1500 data set where the decrease in Sharpe ratio is only about 5%. The Sharpe ratios are similar across the considered dynamic covariance matrix models. The dynamic models outperform the static model for the data sets with more than 10 assets. The gains in out-of-sample Sharpe ratio with the MHEx approach over the static model are about 10% for the FF100 and SP1500 data sets and over 20% for the SP500 data set. Specifically, for the SP500 data set, the annualized Sharpe ratios of 0.90 for MHEx and 0.73 for the static model are significantly different at the 5% level. In addition, the MHEx model yields a statistically significant outperformance of the static model at the 5% level for the FF100 data set. Otherwise, the differences in Sharpe ratios between the covariance matrix models are insignificant. Among the dynamic models, the MHEx model performs best (while the DCC-NL model is second best), which is primarily based on the FF100 stock characteristic portfolios and the single stock SP data sets.

As we do not report the weight statistics of the restricted portfolios in a separate table to save space, we briefly discuss the effects of the portfolio constraints. The statistics for the unconstrained portfolios in Table 4.5 show an average gross leverage above 1.6 for all optimized portfolios. The leverage constraint is thus almost always binding for the constrained 130/30 portfolios. Similarly, the average maximum portfolio weights for the constrained portfolios are at the upper bounds specified for the portfolio weights. The average minimum weights decrease in absolute value due to the leverage constraint and are always greater than the specified lower bounds. The average proportion of negative portfolio weights, i.e., the number of short positions, is greatly reduced by the leverage constraint, from about 50% for the unrestricted portfolios to less than 10% for the restricted portfolios for the FF100 and SP data sets, and to 15–30% for the FF10 and FF30 data sets. Likewise, portfolio turnover is reduced as a result of the portfolio constraints, especially for the mean-variance optimized portfolios. The average turnover decreases by 40–60% for the estimated minimum variance portfolios and by 65–85% for the mean-variance optimal portfolios. For example, for the SP1500 data set, the average monthly turnover with the MHEx model is 1.69 (2.14) for the weight-constrained portfolio instead of 3.03 (13.00) for the unconstrained estimated minimum variance (mean-variance optimal) portfolio. These statistics demonstrate that the practical implementability of the portfolio strategies is greatly enhanced by the introduction of portfolio constraints. A moderate overall decrease in portfolio performance is the result, but using the MHEx model remains beneficial in this more practical setting.

Figure 4.3 shows the cumulative out-of-sample performance paths from daily realized returns of the restricted minimum variance (MINV-MHEx) and mean-variance optimized portfolios (MV-MHEx) based on the MHEx model for the FF100 and SP1500 data sets. We compare their performance to the equally-weighted portfolio, which is a less sophisticated portfolio strategy but performs well in empirical comparisons, see DeMiguel et al. (2009). The estimated minimum variance portfolios show the lowest volatility in both charts and stable performance over time. For the FF100 portfolio data set, the estimated minimum variance portfolio consistently outperforms the equally-weighted portfolio over the 53-year out-of-sample period, and is outperformed by the mean-variance optimized portfolio, which has only slightly higher volatility. For the

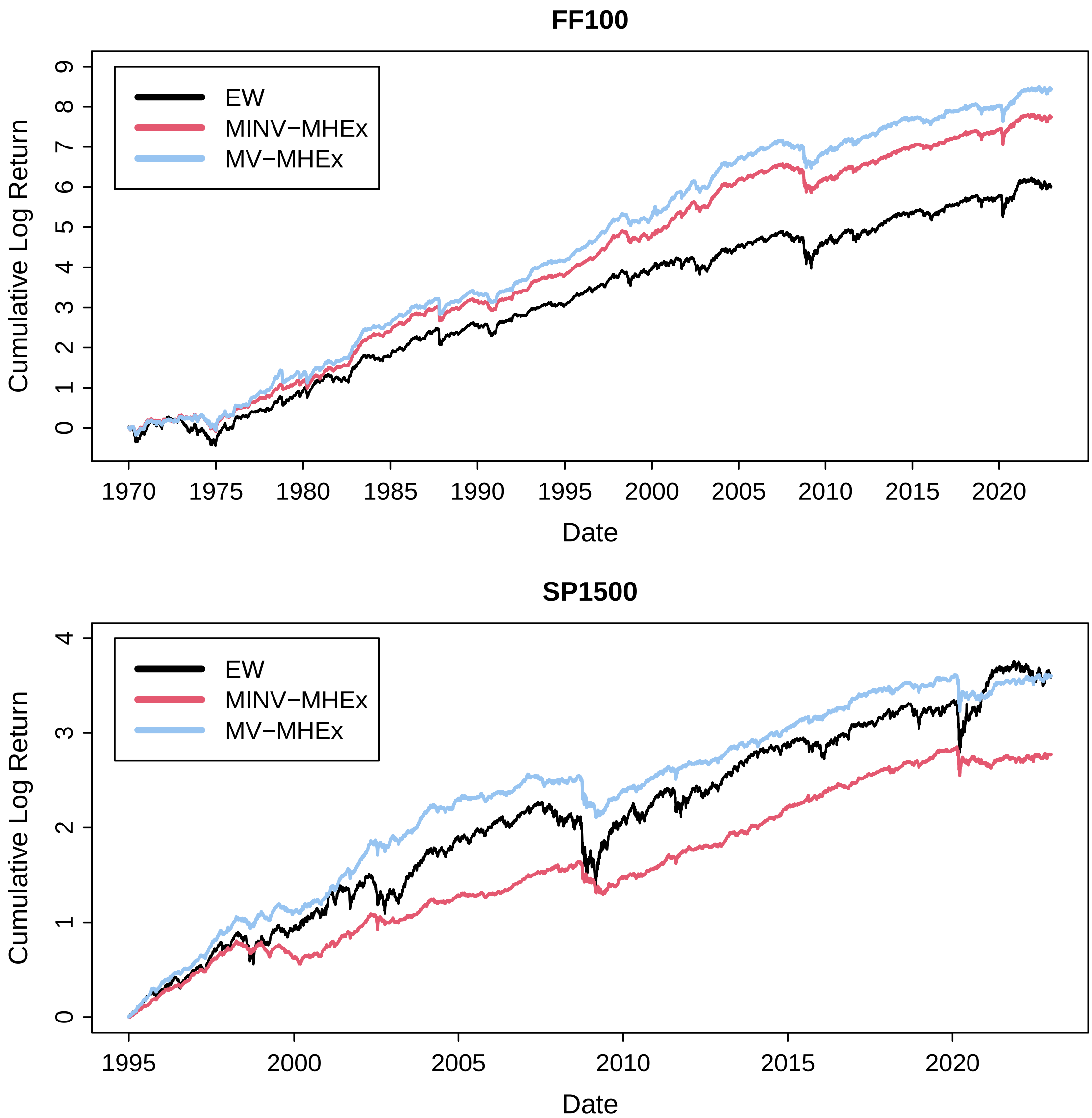

Figure 4.3: Cumulative out-of-sample performance paths based on daily log returns of weight-constrained estimated minimum variance and mean-variance optimal portfolios. "EW" is the equally-weighted portfolio. "MINV-MHEx" and "MV-MHEx" refer to the minimum variance and mean-variance portfolios based on the MHEx model.

SP1500 single stock data set, the estimated minimum variance portfolio ends up below the equally-weighted portfolio, but achieves its goal of producing a low volatility strategy with lower drawdown during periods of distress. The estimated mean-variance optimal portfolio and the equally-weighted portfolio end up at the same point at the end of the 28-year out-of-sample period in 2022. However, the performance path of MV-MHEx is visibly less volatile, implying superior risk-adjusted performance, e.g., in terms of Sharpe ratio, and it has lower drawdown in periods of distress compared to the equally-weighted portfolio.

4.4 Conclusion

In this paper, we propose a framework for modeling large dynamic covariance matrices based on heterogeneous autoregressive volatility and correlation components that have different short- and long-term horizons. Within the framework, called MHEx for short, we incorporate nonlinear shrinkage estimators into the model for correlations, thus effectively addressing estimation risks faced in high-dimensional applications. Because the volatility and correlation components are pre-specified, our model can be estimated via least squares methods. Taken together, these features make the MHEx covariance matrix model flexible, simple to implement, and easily applicable in high-dimensional settings.

We confirm the benefits of the proposed MHEx model through an empirical backtesting exercise. We use the model in monthly portfolio selection and consider asset universes ranging from ten to 1500, where the assets are predefined portfolios or individual stocks. We find that minimum variance portfolios based on covariance matrix forecasts from the MHEx model deliver favorable out-of-sample performance statistics compared to the (R)DCC-NL benchmark models. Moreover, the differences in portfolio standard deviations are statistically significant and economically meaningful on all data sets. For the 100 portfolios formed on size and book-to-market, the annualized portfolio standard deviation improves from 9.49% (9.27%) using the (R)DCC-NL model to 8.62% using our MHEx model. For mean-variance portfolio selection, the optimized portfolios using the MHEx model significantly outperform the equally-weighted benchmark portfolio across all data sets, improving Sharpe ratios by 67 to 114% when weight and leverage constraints are considered. Because unconstrained optimized portfolios often exhibit large portfolio positions, high portfolio leverage, and high turnover, we consider a more practical setting with weight-constrained 130/30 (long-short) minimum variance portfolios. The differences in out-of-sample standard deviations remain statistically significant and economically meaningful. For example, considering the universe of S&P 1500 constituents, the annualized portfolio standard deviation improves to 8.92% using our MHEx model compared to 10.43% (9.75%) using the (R)DCC-NL model.

Our paper offers various directions for future research. To further enhance its forecasting ability, our model could be extended to incorporate semi-(co)variance and realized measures based on high-frequency intraday data. For example, Patton and Sheppard (2015) use intraday-based semi-variances in a HAR model, Bollerslev et al. (2020) use intraday-based semi-covariances in a DCC model, and De Nard et al. (2022) use range-based variance estimators in a DCC-NL model. In addition, for larger stock based data sets, more sophisticated panel methods could be used for model specification and estimation, such as in Qiu et al. (2022) or Wu and Xu (2022).

Chapter 5

Conclusion

Parameter estimation risk is a key reason why applications of portfolio theory do not deliver the promised performance in practice. Estimation risk arises from a lack of knowledge of the true parameters of a model that describes the future behavior of specific variables, e.g., the returns on financial assets. In this circumstance, inferring the parameter values from empirical observations is often the best viable option. However, because, in practice, historical data is limited and may not perfectly represent the actual future behavior, estimated parameters will inevitably deviate from the true model parameters. In the context of portfolio selection, estimation risk causes the estimated optimal portfolio weights to be suboptimal with respect to the true properties of future asset returns. As a consequence, realized portfolio returns will, on average, be lower and portfolio risk will be higher than predicted by portfolio theory. In practice, optimized portfolios often underperform simple portfolio allocation strategies. Thus, it is essential to take estimation risk into account in optimal portfolio selection, specifically within the mean-variance framework as it is widely used in academia and practice.

This dissertation contributes to a better understanding of parameter estimation risks in optimal portfolio selection and provides advanced methods for dealing with estimation risks to effectively improve the practical performance of optimized portfolios. The dissertation consists of three original research papers, each with a specific focus. The first paper in Chapter 2 studies how to estimate mean-variance optimal portfolios with improved out-of-sample performance under a Sharpe ratio criterion using typical input components such as monthly returns and standard sample estimators. Because the estimated optimal portfolio weights are exposed to estimation risk, the paper considers shrinking the in-sample Sharpe ratio optimal portfolio towards the minimum variance portfolio, as this exhibits lower estimation errors. The paper develops a general method based on a bootstrap approach to determine the optimal shrinkage intensity, i.e., how much the estimated portfolio should be shifted towards the estimated minimum variance portfolio. The effectiveness of the proposed method is demonstrated in a simulation study and empirical out-of-sample analysis. In addition to constructing portfolios that take into account estimation risks, the framework incorporates transaction costs incurred in rebalancing the portfolio from the preceding period. The empirical analysis shows that the resulting mean-variance optimized portfolios outperform several benchmark strategies such as the equally-weighted portfolio and the minimum variance portfolio in terms of out-of-sample Sharpe ratio net of transaction costs. The results hold for different datasets representing a variety of investment opportunities.

The second paper in Chapter 3 studies how additional variables can benefit the construction of large mean-variance optimal portfolios of individual stocks given an arbitrary bound on portfolio risk. An empirical application to three broad US and European stock market indices demonstrates that there is significant economic value in specifying a structural model for predicting expected stock returns using firm-specific characteristics. The resulting mean-variance optimal portfolios outperform several benchmark strategies, e.g., the equally-weighted market portfolio or alternative optimized portfolios based on firm characteristics, in terms of out-of-sample Sharpe ratios. To estimate the relationship between firm characteristics and expected returns, the paper considers several machine learning methods. The methods, such as boosted regression trees and artificial neural networks, provide more precise estimates of expected returns by accounting for nonlinear and interaction effects, and yield modest improvements in portfolio performance compared to a linear model. While a linear model already captures much of the main effects, a more refined estimation of expected returns remains inherently difficult due to the low signal-to-noise ratio of individual stock returns. The mean-variance optimized portfolios, taking into account parameter estimation risks, perform well empirically. To avoid excessive portfolio turnover, strategies are implemented that partially rebalance the weights of the optimized portfolios. The results of the empirical analysis show that portfolio turnover can be substantially reduced while maintaining overall favorable portfolio performance.

The third paper in Chapter 4 studies how to model the dynamics of time-varying covariance matrices for a large number of assets, such as individual stocks, in a simple and effective way. The presented framework, called MHEx for short, incorporates separate realized volatility and correlation dynamics using a heterogeneous autoregressive modeling approach. To address parameter estimation risk and ensure that the proposed model performs well in high dimensions, the approach uses shrinkage estimators for the correlation matrices. Because the MHEx model is built based on multiple pre-specified volatility and correlation components that reflect various investor horizons, the model is parsimonious, flexible, and simple to implement. The proposed MHEx model produces direct monthly covariance matrix forecasts using higher frequency daily data and is applicable in high dimensions with more than a thousand assets. In an empirical application to various investment universes ranging in size from ten to 1500 assets, the MHEx model outperforms common benchmark models such as a static covariance matrix model, the

DCC-NL model from Engle et al. (2019), and a Realized DCC model based on Bollerslev et al. (2020). The empirical analysis shows that a risk-averse mean-variance investor benefits from using the MHEx model since it leads to reduced portfolio risk, which is measured by the out-of-sample volatility of realized portfolio returns. The results are robust to different portfolio sizes and asset universes comprising predefined equity portfolios or large numbers of individual stocks, different subperiods of the examined out-of-sample period, different specifications of the considered covariance matrix models, and the introduction of realistic portfolio weight constraints to reflect a more practical setting.

The findings of this dissertation have practical implications for applications of portfolio theory. While portfolio theory states how investors should optimally select their portfolios given assumptions about their preferences, it proves difficult to put into practice under real-world conditions. The dissertation contributes to the understanding of how the theory can be used to improve investors' portfolio choices. Such understanding allows for more effective portfolio management, leading to lower risk with higher return. In general, this is relevant for both private and institutional investors. The methods in Chapters 2 and 4 can be used by sophisticated retail investors for small to medium-sized portfolios, while the applications to larger portfolios and the methods in Chapter 3 are more feasible for institutional investors. In that sense, the dissertation typically takes the perspective of large investors, e.g., banks, insurance companies, investment or pension funds. However, these institutional investors act as intermediaries who manage and invest client funds and therefore can offer more attractive risk-return profiles to the benefit of their clients, when recognizing the investment benefits of efficient portfolio diversification. Moreover, efficient portfolio diversification contributes to improved financial risk management, which aims to mitigate, control, and hedge against risk while maintaining profit potential. At a higher level, risk management is important for the stability of banking and insurance systems (see, e.g., Laeven et al. (2016), Adrian and Brunnermeier (2016), Acharya et al. (2017), and Brownlees and Engle (2017)). At the individual level, appropriate financial risk management protects against bankruptcy and increases the economic value of a company (see, e.g., McNeil et al. (2015, pp. 31–33), Pérez-González and Yun (2013), Ellul and Yerramilli (2013), and Grace et al. (2015)). In Chapters 2–4, this dissertation provides a variety of tools for more accurate measurement and enhanced analysis of risk and return.

Based on these contributions and implications, there is room for additional analysis and further research beyond the scope of this dissertation. The dissertation considers the popular mean-variance framework, which typically works well to approximate the main relationships and effects for optimal portfolio selection in a realistic setting, see Levy and Markowitz (1979), Levy and Levy (2004), and Markowitz (2014). However, the actual preferences and perception of risk among investors can be more complex than implied by mean-variance theory (see the review of Starmer (2000)). For example, attitudes toward risk can differ for gains and losses relative to a reference point. Investors typically give greater weight to losses, especially large losses, than to equivalent gains. This implies they are averse to negative skewness of the portfolio return distribution, i.e., when large negative portfolio returns occur more frequently than large positive returns. Investor preferences of this kind can be modeled by replacing the variance with risk measures that reflect downside risk, such as semivariance (see Markowitz et al. (2020)), value-at-risk (see Gaivoronski and Pflug (2005)), or expected shortfall (see Rockafellar and Uryasev (2000, 2002)), or by using alternative preference frameworks, such as generalized disappointment aversion (see Gul (1991) and Routledge and Zin (2010)) or prospect theory (see Kahneman and Tversky (1979) and Tversky and Kahneman (1992)). The methods and analyses in Chapters 2–4 could be extended to such frameworks taking mean-variance analysis, which includes the variance as a symmetric risk measure, as a useful starting point.

Further research could be conducted related to the approaches in Chapters 2–4. The methods in Chapter 4 can be combined with the approaches in Chapters 2 and 3. That is, the bootstrap approach in Chapter 2 could be extended to incorporate time-varying moments, in particular time-varying second moments, i.e., variances and covariances. For this purpose, one could consider a block bootstrap approach suitable for time series, as proposed in Politis and Romano (1994). Advanced distributional assumptions could be incorporated, allowing for skewness and kurtosis in the multivariate return distribution, such as in a multivariate skewed t-distribution. To estimate the monthly covariance matrix, daily data could be used to take into account higher frequency information as in Chapter 4. The approach in Chapter 3 could also be extended to allow for time-varying covariance matrices given the methodology in Chapter 4. In addition, a wider range of firm characteristics and other predictive variables, such as option-implied variables, could be implemented for the expected return model, and further analysis could be conducted, e.g., by including small firms not represented in the market indices and by extending the time series by utilizing additional data sources. It could be interesting to analyze the asset pricing implications, e.g., along the lines of DeMiguel et al. (2020) or Daniel et al. (2020), that result from using the proposed methodology.

With respect to the methods in Chapter 4, alternative shrinkage estimators of the covariance matrix could be developed specifically for the case where the number of assets exceeds the number of observations and for exponential weighting of observations. The presented MHEx model can be easily used to predict covariances at frequencies other than the monthly frequency. At the daily frequency, the model could be estimated using a likelihood-based approach and compared to the popular univariate GARCH model or the multivariate DCC model. In addition, starting from a univariate setting, it could be interesting to extend the heterogeneous autoregressive approach to model time-varying higher moments such as skewness, e.g., within a skewed t

distribution. In a multivariate setting, the modeling framework could be incorporated into a copula model that allows for both asymmetric and tail dependence, which are commonly observed empirical features of financial asset returns, see, e.g., Christoffersen and Langlois (2013) and Christoffersen et al. (2012, 2018).

With new technologies emerging in the future, for example in the fields of machine learning and artificial intelligence, and their increasingly refined application, we can expect further enhancements for applications of portfolio theory and the methods for dealing with parameter estimation risks.

References

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2017). Measuring systemic risk. *The Review of Financial Studies 30*(1), 2–47.

Adrian, T. and M. K. Brunnermeier (2016). CoVaR. *American Economic Review 106*(7), 1705–1741.

Alves, R. P., D. S. de Brito, M. C. Medeiros, and R. M. Ribeiro (2023). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage. *Journal of Financial Econometrics (Forthcoming)*, https://doi.org/10.1093/jjfinec/nbad013.

Ammann, M., G. Coqueret, and J.-P. Schade (2016). Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance 70*, 23–37.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The Journal of Finance 61*(1), 259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics 91*(1), 1–23.

Ao, M., L. Yingying, and X. Zheng (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies 32*(7), 2890–2919.

Bali, T. G. and N. Cakici (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis 43*(1), 29–58.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Ban, G.-Y., N. El Karoui, and A. E. B. Lim (2018). Machine learning and portfolio optimization. *Management Science 64*(3), 1136–1154.

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

Barbee, W. C., S. Mukherji, and G. A. Raines (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal 52*(2), 56–60.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Basak, G. K., R. Jagannathan, and T. Ma (2009). Jackknife estimator for tracking error variance of optimal portfolios. *Management Science 55*(6), 990–1002.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance 32*(3), 663–682.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics 12*(1), 129–156.

Bauer, G. H. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics 160*(1), 93–101.

Bauwens, L. and Y. Xu (2023). DCC- and DECO-HEAVY: Multivariate GARCH models based on realized variances and correlations. *International Journal of Forecasting 39*(2), 938–955.

BCG (2023). Global asset management 2023—21st annual report: The tide has turned. Available at: https://www.bcg.com/publications/2023/the-tide-has-changed-for-asset-managers.

Behr, P., A. Guettler, and F. Miebs (2013). On portfolio optimization: Imposing the right constraints. *Journal of Banking & Finance 37*(4), 1232–1242.

Bekaert, G., E. C. Engstrom, and N. R. Xu (2022). The time variation in risk appetite and uncertainty. *Management Science 68*(6), 3975–4004.

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. *The Review of Financial Studies 4*(2), 315–342.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance 43*(2), 507–528.

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

Bloomfield, T., R. Leftwich, and J. B. Long (1977). Portfolio strategies and performance. *Journal of Financial Economics 5*(2), 201–218.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies 31*(7), 2729–2773.

Bollerslev, T., N. Meddahi, and S. Nyawa (2019). High-dimensional multivariate realized volatility estimation. *Journal of Econometrics 212*(1), 116–136.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2020). Multivariate leverage effects and realized semicovariance GARCH models. *Journal of Econometrics 217*(2), 411–430.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2022). Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics 144*(1), 227–246.

Brailsford, T. J. and R. W. Faff (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance 20*(3), 419–438.

Brandt, M. W. (2010). Portfolio choice problems. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and Techniques*, pp. 269–336. Amsterdam: North Holland/Elsevier.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies 22*(9), 3411–3447.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research 45*(1), 21–58.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences 106*(30), 12267–12272.

Brownlees, C. and R. F. Engle (2017). SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies 30*(1), 48–79.

Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. Working paper, https://doi.org/10.2139/ssrn.3493458.

Chan, L. K. C., J. Karceski, and J. Lakonishok (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies 12*(5), 937–974.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Chen, T. and C. Guestrin (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, pp. 785–794.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26*(6), 922–947.

Christoffersen, P., V. Errunza, K. Jacobs, and H. Langlois (2012). Is the potential for international diversification disappearing? A dynamic copula approach. *The Review of Financial Studies 25*(12), 3711–3751.

Christoffersen, P., K. Jacobs, X. Jin, and H. Langlois (2018). Dynamic dependence and diversification in corporate credit. *Review of Finance 22*(2), 521–560.

Christoffersen, P. and H. Langlois (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis 48*(5), 1371–1404.

Cochrane, J. H. (2005). *Asset Pricing*. Princeton: Princeton University Press.

Cohen, K. J. and J. A. Pogue (1967). An empirical evaluation of alternative portfolio-selection models. *The Journal of Business 40*(2), 166–193.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance 63*(4), 1609–1651.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Creal, D. D. and R. S. Tsay (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics 189*(2), 335–345.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020). The cross-section of risk and returns. *The Review of Financial Studies 33*(5), 1927–1979.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

De Nard, G., R. F. Engle, O. Ledoit, and M. Wolf (2022). Large dynamic covariance matrices: Enhancements based on intraday data. *Journal of Banking & Finance 138*, 106426.

De Nard, G., O. Ledoit, and M. Wolf (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics 19*(2), 236–257.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science 55*(5), 798–812.

DeMiguel, 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.

DeMiguel, V., A. Martin-Utrera, and F. J. Nogales (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance 37*(8), 3018–3034.

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies 33*(5), 2180–2222.

DeMiguel, V. and F. J. Nogales (2009). Portfolio selection with robust estimation. *Operations Research 57*(3), 560–577.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis 48*(6), 1813–1845.

Dickinson, J. P. (1974). The reliability of estimation procedures in portfolio analysis. *Journal of Financial and Quantitative Analysis 9*(3), 447–462.

Ellul, A. and V. Yerramilli (2013). Stronger risk controls, lower risk: Evidence from U.S. bank holding companies. *The Journal of Finance 68*(5), 1757–1803.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*(2), 363–375.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *The Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (2006). The value premium and the CAPM. *The Journal of Finance 61*(5), 2163–2185.

Fama, E. F. and K. R. French (2008). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Frankfurter, G. M., H. E. Phillips, and J. P. Seagle (1971). Portfolio selection: The effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis 6*(5), 1251–1262.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics 19*(1), 3–29.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Fried, J. (1970). Forecasting and probability distributions for models of portfolio selection. *The Journal of Finance 25*(3), 539–554.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics 29*(5), 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis 38*(4), 367–378.

Friend, I. and D. Vickers (1965). Portfolio selection and investment performance. *The Journal of Finance 20*(3), 391–415.

Frost, P. A. and J. E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis 21*(3), 293–305.

Frost, P. A. and J. E. Savarino (1988). For better performance: Constrain portfolio weights. *The Journal of Portfolio Management 15*(1), 29–34.

Füss, R., F. Miebs, and F. Trübenbach (2014). A jackknife-type estimator for portfolio revision. *Journal of Banking & Finance 43*, 14–28.

Gaivoronski, A. A. and G. Pflug (2005). Value-at-risk in portfolio optimization: Properties and computational approach. *Journal of Risk 7*(2), 1–31.

Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies 20*(1), 41–81.

Ghysels, E., A. Plazzi, R. Valkanov, A. Rubia, and A. Dossani (2019). Direct versus iterated multiperiod volatility forecasts. *Annual Review of Financial Economics 11*(1), 173–195.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. Cambridge: MIT Press.

Gourieroux, C., J. Jasiak, and R. Sufana (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181.

Grace, M. F., J. T. Leverty, R. D. Phillips, and P. Shimpi (2015). The value of investing in enterprise risk management. *Journal of Risk and Insurance 82*(2), 289–316.

Graham, B. and D. L. Dodd (1934). *Security Analysis*. New York: McGraw-Hill.

Gribisch, B. and J. P. Hartkopf (2023). Modeling realized covariance measures with heterogeneous liquidity: A generalized matrix-variate Wishart state-space model. *Journal of Econometrics 235*(1), 43–64.

Gribisch, B., J. P. Hartkopf, and R. Liesenfeld (2020). Factor state–space models for high-dimensional realized covariance matrices of asset returns. *Journal of Empirical Finance 55*, 1–20.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1, Part B), 429–450.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica 59*(3), 667–686.

Hand, J. R. M. and J. Green (2011). The importance of accounting information in portfolio optimization. *Journal of Accounting, Auditing & Finance 26*(1), 1–34.

Hanna, J. D. and M. J. Ready (2005). Profitable predictability in the cross section of stock returns. *Journal of Financial Economics 78*(3), 463–505.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. New York: Springer-Verlag.

Haugen, R. A. and N. L. Baker (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics 41*(3), 401–439.

Hjalmarsson, E. and P. Manchev (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance 36*(5), 1392–1401.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

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

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance 56*(2), 699–720.

Jobson, J. D. and B. Korkie (1981). Putting Markowitz theory to work. *The Journal of Portfolio Management 7*(4), 70–74.

Jobson, J. D., B. Korkie, and V. Ratti (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. In *Proceedings of the American Statistical Association (Business and Economic Statistics Section)*, pp. 279–284.

Jorion, P. (1985). International portfolio diversification with estimation risk. *The Journal of Business 58*(3), 259–278.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis 21*(3), 279–292.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Kalymon, B. A. (1971). Estimation risk in the portfolio selection model. *Journal of Financial and Quantitative Analysis 6*(1), 559–582.

Kan, R. and D. R. Smith (2008). The distribution of the sample minimum-variance frontier. *Management Science 54*(7), 1364–1380.

Kan, R., X. Wang, and G. Zhou (2022). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science 68*(3), 2047–2068.

Kan, R. and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis 42*(3), 621–656.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 222*(3), 501–524.

Kim, S., R. A. Korajczyk, and A. Neuhierl (2021). Arbitrage portfolios. *The Review of Financial Studies 34*(6), 2813–2856.

Kingma, D. P. and J. Ba (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, San Diego, USA.

Kirby, C. and B. Ostdiek (2012). It's all in the timing: Simple active portfolio strategies that outperform naïve diversification. *Journal of Financial and Quantitative Analysis 47*(2), 437–467.

Kourtis, A. (2015). A stability approach to mean-variance optimization. *Financial Review 50*(3), 301–330.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Kritzman, M., S. Page, and D. Turkington (2010). In defense of optimization: The fallacy of $1/N$. *Financial Analysts Journal 66*(2), 31–39.

Kubokawa, T., C. P. Robert, and A. K. M. E. Saleh (1993). Estimation of noncentrality parameters. *Canadian Journal of Statistics 21*(1), 45–57.

Laeven, L., L. Ratnovski, and H. Tong (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking & Finance 69 (Supplement 1)*, S25–S34.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance 49*(5), 1541–1578.

Leavens, D. H. (1945). Diversification of investments. *Trusts and Estates 80*(May), 469–473.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10*(5), 603–621.

Ledoit, O. and M. Wolf (2004a). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O. and M. Wolf (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*(2), 365–411.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Ledoit, O. and M. Wolf (2011). Robust performance hypothesis testing with the variance. *Wilmott 2011*(55), 86–89.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Ledoit, O. and M. Wolf (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies 30*(12), 4349–4388.

Ledoit, O. and M. Wolf (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics 48*(5), 3043–3065.

Ledoit, O. and M. Wolf (2022). Quadratic shrinkage for large covariance matrices. *Bernoulli 28*(3), 1519–1547.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies 33*(5), 2274–2325.

Levy, H. and M. Levy (2004). Prospect theory and mean-variance analysis. *The Review of Financial Studies 17*(4), 1015–1041.

Levy, H. and H. M. Markowitz (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review 69*(3), 308–317.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review 4*(1), 1–44.

Litzenberger, R. H. and K. Ramaswamy (1982). The effects of dividends on common stock prices: Tax effects or information effects? *The Journal of Finance 37*(2), 429–443.

Lo, A. W. and P. N. Patel (2008). 130/30: The new long-only. *The Journal of Portfolio Management 34*(2), 12–38.

Mao, J. C. T. and C. E. Särndal (1966). A decision theory approach to portfolio selection. *Management Science 12*(8), 323–333.

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

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

Markowitz, H. (2014). Mean-variance approximations to expected utility. *European Journal of Operational Research 234*(2), 346–355.

Markowitz, H. M. (1999). The early history of portfolio theory: 1600–1960. *Financial Analysts Journal 55*(4), 5–16.

Markowitz, H. M., D. Starer, H. Fram, and S. Gerber (2020). Avoiding the downside: A practical review of the critical line algorithm for mean-semivariance portfolio optimization. In J. B. Guerard and W. T. Ziemba (Eds.), *Handbook of Applied Investment Research*, Volume 9 of *World Scientific Handbook in Financial Economics Series*, pp. 369–415. New Jersey: World Scientific.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative Risk Management: Concepts, Techniques and Tools—Revised Edition*. Princeton: Princeton University Press.

Mei, X., V. DeMiguel, and F. J. Nogales (2016). Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance 69*, 108–120.

Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics 8*(4), 323–361.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal 45*(1), 31–42.

Moreira, A. and T. Muir (2017). Volatility-managed portfolios. *The Journal of Finance 72*(4), 1611–1644.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working paper, https://doi.org/10.2139/ssrn.2740751.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Oh, D. H. and A. J. Patton (2023). Dynamic factor copula models with estimated cluster assignments. *Journal of Econometrics 237*(2, Part C), 105374.

Okhrin, Y. and W. Schmid (2006). Distributional properties of portfolio weights. *Journal of Econometrics 134*(1), 235–256.

Olivares-Nadal, A. V. and V. DeMiguel (2018). Technical note—A robust perspective on transaction costs in portfolio optimization. *Operations Research 66*(3), 733–739.

Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics 36*(4), 643–657.

Opschoor, A., A. Lucas, I. Barra, and D. van Dijk (2021). Closed-form multi-factor copula models with observation-driven dynamic factor loadings. *Journal of Business & Economic Statistics 39*(4), 1066–1079.

Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2021). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics 39*(3), 652–668.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics 104*(1), 162–185.

Patton, A. J. and K. Sheppard (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics 97*(3), 683–697.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Pontiff, J. and A. Woodgate (2008). Share issuance and cross-sectional returns. *The Journal of Finance 63*(2), 921–945.

Pérez-González, F. and H. Yun (2013). Risk management and firm value: Evidence from weather derivatives. *The Journal of Finance 68*(5), 2143–2176.

Qiu, Y., T. Xie, J. Yu, and Q. Zhou (2022). Forecasting equity index volatility by measuring the linkage among component stocks. *Journal of Financial Econometrics 20*(1), 160–186.

Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk 2*(3), 21–42.

Rockafellar, R. T. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance 26*(7), 1443–1471.

Rosenberg, B., K. Reid, and R. Lanstein (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management 11*(3), 9–16.

Routledge, B. R. and S. E. Zin (2010). Generalized disappointment aversion and asset prices. *The Journal of Finance 65*(4), 1303–1332.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), 425–442.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature 38*(2), 332–382.

Tu, J. and G. Zhou (2010). Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis 45*(4), 959–986.

Tu, J. and G. Zhou (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics 99*(1), 204–215.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Williams, J. B. (1938). *The Theory of Investment Value*. Cambridge: Harvard University Press.

Wu, L. and Y. Xu (2022). Predicting stock return variance in a large cross section. Working paper, https://doi.org/10.2139/ssrn.4171219.

Zhou, G. (2009). Beyond Black–Litterman: Letting the data speak. *The Journal of Portfolio Management 36*(1), 36–45.