Dynamic Modeling and Forecasting of Financial Portfolio Weights

Inauguraldissertation zur Erlangung des Doktorgrades der

Wirtschafts- und Sozialwissenschaftlichen Fakultät der Unversität zu Köln

2022

vorgelegt von

Laura Annabelle Reh, M.Sc.,

aus

Trier

Für alle Menschen, die mich immer unterstützen, und Hera.

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Institut für Ökonometrie und Statistik der Universität zu Köln. Mein ganz besonderer Dank gilt meinem Doktorvater Professor Dr. Roman Liesenfeld. Die gute Zusammenarbeit sowie seine Betreuung und große Unterstützung haben entschieden zum Gelingen dieser Arbeit beitragen. Herrn Professor Dr. Breitung danke ich vielmals für die Übernahme des Zweitgutachtens. Darüber hinaus möchte ich Herrn Juniorprofessor Dr. Fabian Krüger und Herrn Professor Guilherme Valle Moura danken, die neben Herrn Professor Dr. Roman Liesenfeld und mir ebenfalls an den dieser Arbeit zugrundeliegenden Papieren mitgewirkt haben.

Mein großer Dank gilt auch allen (ehemaligen) ArbeitskollegInnen am Institut für Ökonometrie und Statistik, welches mich nicht nur während der Promotion, sondern auch schon beinahe meine gesamte Studienzeit über begleitet hat. Ein hohes Maß an Kollegialität und Hilfsbereitschaft untereinander hat für ein angenehmes und konstruktives Arbeitsklima und starken Zusammenhalt gesorgt, selbst in Zeiten der Pandemie. Ganz besonders hervorheben möchte ich Herrn Dr. Jan Patrick Hartkopf, Herrn Dr. Simon Umbach und Herrn Ilya Zarubin, MSc, die Zeit mit ihnen bleibt unvergesslich. Nicht zuletzt bedanke ich mich bei meinen FreundInnen und meiner Familie, insbesondere meinen Eltern. Sie haben mir immer den Rückhalt gegeben, der mich (nicht nur) durch die letzten Jahre getragen hat. Ohne sie hätte ich die Arbeit nicht schreiben können.

Köln, am 28. Januar 2022   Laura Annabelle Reh

Contents

Contents

List of Tables

List of Figures

Chapter 1.

Introduction

This thesis addresses the modeling and prediction of portfolio weights in high-dimensional applications to returns on a set of risky financial assets with a particular emphasis on Global Minimum Variance Portfolio (GMVP) allocations. A key aspect of active portfolio management is the projection of weights that optimize portfolio holdings with respect to a representative measure of an investor's preferences. Since the seminal work of Markowitz (1952), such projections of optimal portfolio weights have generally been derived from projections of the first two moments of asset returns. The set of mean-variance efficient portfolios contains all allocations that a rational, risk-averse investor would choose. They provide the highest expected portfolio return for a given level of risk chosen according to the investor's risk preferences. This includes the global minimum variance portfolio, which is chosen for the special case of an infinitely risk-averse investor as well as under the assumption that expected returns are the same for all assets.

The global minimum variance portfolio allocates a given budget to n financial assets in such a way that the variance and thus the risk of the portfolio return is minimized. Thus, opposite to all other mean-variance optimal strategies, the weights of the GMVP are solely a function of the (co)variation among the asset returns and do not explicitly depend on their expectations. These, otherwise, have a large impact on the allocation (Best and Grauer, 1991), but are notoriously difficult to predict (Welch and Goyal, 2008), which has been shown to lead to high estimation errors associated with poor out-of-sample performance (Jobson and Korkie, 1980, 1981). Several studies present evidence that the GMVP allocation generally has better out-of-sample performance than other mean-variance efficient strategies, even when performance measures that depend on average returns are considered (see e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009). As a result, GMVP is one of the most widely used investment strategies by both practitioners and researchers in finance.

Due to better data availability and higher computational power, modern asset allocation usually considers hundreds of financial assets. The econometric problem, especially when following the GMVP strategy, is to model the covariation between asset returns. Specifically, classical approaches rely on a plug-in approach to calculate the portfolio weights for a given estimate of the joint covariance matrix Σ of asset returns. However, the number of elements in Σ increases quadratically with the number of assets. This leads to critical

concentration ratios, i.e., situations where the number of cross-sectional units is high relative to the number of estimation periods, which implies that these approaches are prone to the curse of dimensionality. This problem, which makes inference of covariances difficult both numerically and statistically, affects estimates of portfolio weights as well. The challenges are even more pronounced when using dynamic models that predict conditional covariances while accounting for the potential conditional heteroscedasticity that financial time series typically exhibit.

To avoid substantial estimation errors, it is helpful to constrain the covariance: in addition to sparsity assumptions or imposing factor structures, shrinkage techniques (Ledoit and Wolf, 2003, 2004, 2012, 2015) have proven to be very effective in reducing estimation noise. A simple way to incorporate shrinkage into dynamic modeling approaches was proposed by Engle et al. (2019), who use shrinkage estimates of the unconditional correlation matrix as the target matrix in the dynamic conditional correlation (DCC) model of Engle (2002). However, models for covariance matrices – even when designed to prevent overfitting – are not directly related to the weight allocation problem that is actually at hand. Hence, there is a risk that the fitted (potentially regularized) model will fail to capture the economically important properties of weights and their dynamics while capturing irrelevant ones. This implies, for example, that even when regularized covariance estimates are used, undesirable or economically implausible allocations cannot be directly prevented.

Based on the argument that the portfolio weights are a scaled linear combination of the entries of the inverse covariance matrix rather than the covariance matrix itself, it is also possible to build a parsimonious model for the precision matrix of the data in order to establish a more direct relationship of the estimation problem to the allocation problem. A recent example is the approach of Callot et al. (2019), which combines LASSO (least-absolute shrinkage and selection operator) estimation with a nodewise regression representation for the elements of the precision matrix. Alternatively, portfolio weights can be modeled directly from return observations, e.g., based on linear regression representations (Britten-Jones, 1999; Kempf and Memmel, 2006), where the population regression coefficients implicitly capture the correlation structure between the returns.

This thesis contributes to the literature by proposing several approaches to dynamic portfolio optimization that differ in their methodological approaches but agree in addressing the challenges discussed previously: All proposed models allow for dynamic evolution of portfolio weights and are specified to allow scalability in terms of the number of cross-sectional units from both a statistical and computational perspective. The models are designed to predict portfolio weights that represent the quantity of interest, resulting in particularly good out-of-sample performance compared to existing benchmarks. The focus is hereby on GMVP allocations. Specifically, the dissertation includes two regression-based approaches to directly infer the weights of the global minimum variance portfolio. While the first approach is a semiparametric loss function-based method that does not rely on distributional

assumptions for the returns, the second defines GMVP regression in a state-space framework, using Bayesian inference that allows for data-driven regularization through the specification of the priors. The third approach presents a time-varying Gaussian graphical model for the precision matrix of the return data. Equipped with adaptive penalties for the graph and its evolution tailored to the allocation problem at hand, the detour via the precision is rewarded by straightforward applicability to a broader class of (potentially constrained) allocation problems than the first two approaches. All of the proposed models are computationally tractable and applicable in large dimensional environments as well as under challenging concentration ratios, making them valuable to practitioners and researchers in the field of finance and financial econometrics.

In total, this thesis comprises three self-contained essays on modeling and predicting dynamic portfolio weights for financial asset returns. The essay in Chapter 2 is a joint paper with Prof. Dr. Roman Liesenfeld and Jun.-Prof. Dr. Fabian Krüger. The paper in Chapter 3 is a joint work with Prof. Dr. Roman Liesenfeld and Prof. Dr. Guilherme Valle Moura. The essay in Chapter 4 is a single authored project. The essays and my contributions to them are summarized below.

Chapter 2 corresponds to the paper 'Predicting the Global Minimum Variance Portfolio' (Reh et al., 2021) which is accepted in its current form for publication in the *Journal of Business & Economic Statistics*. The paper proposes a direct modeling approach for the global minimum variance portfolio (GMVP) weights, without a detour via the covariance or its inverse. We develop a dynamic GMVP approach that accounts for the autoregressive conditional heteroscedasticity of asset returns and aims to predict the conditional GMVP weights, i.e., the weights that minimize the conditional variance of portfolio returns. The inference of the weights is done by minimizing a consistent loss function (Gneiting, 2011) that results from a representation of the GMVP weights as population coefficients in an auxiliary linear regression (Kempf and Memmel, 2006) of a benchmark return on a vector of return differentials. The approach thus requires no distributional assumptions for the returns, which makes it robust to misspecification. In addition to this, we provide a detailed conceptual motivation for a dynamic conditional approach based on this GMVP loss function. In particular, we show that the expected GMVP return that enters the GMVP loss function as a nuisance parameter is unavoidable, i.e., there can be no loss function that uniquely identifies the GMVP weights alone. Moreover, we argue that the expected GMVP loss is usually very similar to the unconditional portfolio variance, which is the key criterion for GMVP portfolio evaluation in practice. At the same time, the use of a loss function (instead of the unconditional variance) allows for tailored model development, parameter estimation, and forecast evaluation using model confidence sets (Hansen et al., 2011) and related tools. Our approach combines a direct dynamic parameterization of the weights with a consistent loss function to estimate the parameters and evaluate the resulting predictions. This consistent setup contrasts with much existing work that uses different loss functions to estimate and

evaluate an econometric model. Finally, motivated by theoretical results of Ferson and Siegel (2001), we show through simulations of artificial data that in realistic scenarios for return data at high sampling frequencies such as daily data with pronounced time variation in the conditional covariance matrix of asset returns and much smaller variation in their conditional expectations, dynamic conditional approaches can be expected to outperform static strategies which do not explicitly account for conditional heteroscedasticity. We then use the GMVP loss function to develop time series models for the conditional GMVP weights, relying on recursive least squares (RLS) with forgetting (Ljung and Söderström, 1983) or generalized autoregressive score (GAS; Creal et al., 2013) recursions. With a large number of assets, a dynamic GMVP model requires parsimonious parameterization to be manageable in practice. In addition, the model should not be influenced by the choice of the base asset in the GMVP loss function. This choice, while necessary, is arbitrary and therefore should not affect the predicted GMVP weights of the model. These two requirements motivate our proposed dynamic GMVP models. Our RLS version of the GMVP model is parsimonious by design, and in the GAS version, parsimony is achieved by specifying a sparse update recursion that targets the equally weighted portfolio. We further show that these RLS and GAS specifications are invariant with respect to the benchmark asset.

To investigate the performance of our proposed models, we first conduct a Monte Carlo experiment, which shows that our proposed specifications are able to track the 'true' conditional GMVP weights and perform very well, especially in high concentration ratio situations. Subsequently, in an empirical application to daily U.S. traded stock returns with a cross-sectional dimension from $n = 50$ to $n = 1000$ and different lengths of the estimation window, we find that they perform very well in terms of both expected loss and unconditional portfolio variance compared to existing static and dynamic approaches. In particular, by using a regularized exponential forgetting version of recursive least-squares dynamics, we regularize the dispersion of the eigenvalues of the RLS information matrix, which makes the model robust to large concentration ratios, even for $n > T$. It is worth mentioning that in two earlier versions of the paper we explored a much more extensive list of model specifications for the weight dynamics, in particular higher parameterized and thus more flexible GAS structures, as well as GAS and RLS models in combination with nonlinear shrinkage estimates of the unconditional GMVP weights. In particular, models with initialization based on nonlinear shrinkage were found to be well applicable for medium to large cross-sectional sizes ($n \leq 200$). In addition, we have shown in previous versions of the paper that our proposed models also perform well relative to benchmark approaches when applied to monthly return data such as industry portfolios or data sets composed based on Fama-French factors.[1]

My contributions to Chapter 2 are as follows: In addition to assisting the writing process by constantly revising drafts of the paper, I contributed to the development of the dynamic GAS and RLS model specifications and collaborated on the design of the simulation study

---

[1] Corresponding model descriptions, additional out-of-sample forecasting results for these models and further notes on the M-estimator for the model parameters are provided in the Appendix for Chapter 2.

and empirical application. I conducted all implementations for the simulations and empirical application, which also includes the implementation of the benchmark models. Moreover, I contributed to the derivations on (conditional) optimality as well as to the verification of the invariance assumption for the proposed model specifications. I also derived analytical gradients, which are, however, no longer included in the current version of the paper, but stated in the Appendix.

Chapter 3 corresponds to the working paper 'A Time-Varying Parameter Model with Bayesian Shrinkage for Global Minimum Variance Portfolio Prediction'. The paper proposes a direct modeling approach for the global minimum variance portfolio (GMVP) weights in the framework of a state-space model. We use Bayesian techniques for the inference of the time-varying weights which allows to impose regularization in a data driven way via the prior choice. An advantage of our approach is that the problem scales linearly, so it is easily applicable to large dimensions and gives good results even if the concentration ratio exceeds one. In particular, shrinkage-type specifications for the priors are very beneficial in such scenarios. Moreover, since we only work with the quantity of interest for portfolio allocation, we can avoid direct estimation of large and often ill-conditioned covariance matrices. Unlike plug-in approaches that use estimates of the covariance matrix to compute GMVP weights, the proposed method allows for direct shrinkage of portfolio weights.

Similar to the approach in Chapter 2, the model is based on a representation of GMVP weights via a linear regression of returns on a reference portfolio on a vector of return differentials, where the error variance equals the variance of the GMVP return, which represents the systemic risk that cannot be further reduced by diversification. The evolution of portfolio weights is determined by the evolution of underlying latent state processes. The introduction of Gaussian random walk dynamics for the parameters and a Gaussian assumption for the error term in the measurement equation combined with stationary AR(1) dynamics for the log volatility process leads to a linear Gaussian state space model with time-varying parameters and stochastic volatility (SV), which can be easily treated using standard Markov Chain Monte Carlo (MCMC) procedures. This specification allows to account for both the time variation in the assets' conditional covariance structure and the heteroscedasticity in the market, which corresponds to the heteroscedasticity of the error term in the measurement equation. In principle, it would be easy to account for potential heavy tails by replacing the normal distribution for the error term in the measurement equation by a Student-t, but preliminary results suggest that this changes the results only slightly.

In contrast to the auxiliary regression employed in Chapter 2, here we adopt the approach of Frey and Pohlmeier (2016) and extend the asset space by one dimension by setting the benchmark return to a linear combination of the returns such as the equally weighted portfolio. In this way, the population regression coefficients can be interpreted as deviations from the considered benchmark portfolio. Imposing hierarchical Bayesian LASSO-type priors on the regression coefficients implies that the portfolio weights are shrunk in the direction of

the reference portfolio, which is, when using the equally weighted portfolio as reference as we propose for our primarily used specifications, equivalent to imposing norm constraints on the portfolio weights. Unlike norm-regularized frequentist estimation, however, imposing hierarchical priors corresponds to probabilistic shrinkage of the parameter space, so that no hard constraints and virtually no calibrations need to be imposed a-priori. One contribution of the paper is that we extend the Bayesian shrinkage approach for portfolio weights to a dynamic context. We formulate the dynamics for the time-varying parameters in so-called non-centered form. This means that we have one additive parameter each for the initial level, as well as one multiplicative parameter for the standard deviation, while the latent state processes themselves are independent of any parameter. As prior distribution for the standard deviation in the random walk dynamics, we choose the Bayesian LASSO specification of Belmonte et al. (2014), a hierarchical Gaussian prior which ensures that the time variation is regularized for all portfolio weights. For the overall level of initial values corresponding to the initial deviation from the reference portfolio, we assign the recently proposed double Gamma prior (Bitto and Frühwirth-Schnatter, 2019), which compared to the specification of Belmonte et al. (2014) adds a hyperprior to the parameter determining the variation in the shrinkage intensities along the parameters, and thus allows for greater local flexibility at a global shrinkage level. Our Bayesian shrinkage priors for the dynamics and initial level of portfolio weights in direction of the benchmark portfolio simultaneously perform two tasks: First, they reduce estimation noise, allowing us to deal with difficult concentration ratios. Second, they regulate allocations, which, particularly when using the naïve portfolio as benchmark, reduces short sales and exposures as well as the volatility of weights, leading to a smoother evolution of weights associated with moderate turnover.

An advantage of the reference augmented formulation of the problem is that it circumvents the choice of a benchmark among the returns considered and ensures that the approach is invariant with respect to the ordering of assets, since all assets are – up to alternative benchmark choices – treated symmetrically by construction. Adding one dimension to the asset space implies that the regressors are collinear. However, the regression coefficients are still uniquely identified up to an additive constant. Therefore, in the framework of our Bayesian analysis, the problem can be cured by a unimodal prior distribution. In our analysis, we ensure the existence of uniquely identified posterior distributions via a normal distribution with finite variance (double Gamma prior) for the prior of the initial level parameters of the regression coefficients.

The applicability and robustness of the approach is demonstrated through extensive simulations and empirical analysis. In particular, a simulation study based on a DCC data generation process shows that the proposed approach can perform better than the true model when the number of observations is not much larger than the number of assets, for both in- and out-of-sample comparisons. An application to daily financial returns for dimensions up to $n = 1600$ also shows that the time-varying parameter GMVP model outperforms

a wide range of existing approaches in terms of out-of-sample forecast accuracy. The application of shrinkage priors is particularly important in scenarios with difficult concentration ratios. Further, it is shown that accounting for stochastic volatility improves the overall results, likely due to the fact that any noise would otherwise be incorrectly accounted for in the dynamics of the parameters and thus the portfolio weights. Remarkably, our GMVP predictions are also found to be a strong competitor in terms of mean-variance efficiency, both without and with transaction costs taken into account. As an alternative to a stationary AR(1) process for the log-volatility dynamics, we initially have also run our model with a random walk dynamics for the log-squared volatility as well as a Beta process for the inverse volatility states following the approach of Uhlig (1994, 1997), which have also yielded satisfactory results. Details on these alternative SV specifications and corresponding out-of-sample results are given in the Appendix for Chapter 3.

My contributions to Chapter 3 are as follows: First, I have developed the idea of formulating the dynamic GMVP problem in an augmented asset space as a linear state-space model to apply Bayesian inference techniques that allow data-driven shrinkage through certain LASSO-type prior specifications. Second, I have collaborated on the selection of the particular prior specifications and the selection of the stochastic volatility dynamics. Third, I fully implemented in MATLAB the proposed MCMC scheme for parameter estimation and prediction as well as the Rao-Blackwellized particle filter that is applied in the in-sample simulation study. The implementation also includes the benchmark models. Fourth, I added the section on model extensions (Section 3.6.4). Finally, I authored a substantial part of the paper and constantly revised the complete draft of the paper.

Chapter 4 corresponds to the working paper 'Inferring dynamic financial networks via a time-varying graphical LASSO approach with applications to portfolio selection', which is currently under review at the *International Journal of Forecasting*. In this work, I propose to use a time-varying graphical LASSO approach to model inverse covariance matrices in a large-dimensional system of financial assets, with the goal of predicting optimal portfolio allocations. In a Gaussian model, the precision matrix can be interpreted as an undirected graph in which the edges correspond to pairwise partial correlations between two return series. By modeling financial precision matrices instead of the covariance, one can exploit that mean-variance optimal portfolio weights are linked to the elements of the precision matrix via scaled linear functions. This facilitates interpretation and allows specific requirements such as regularizing the weights to reduce estimation noise to be addressed during the precision estimation process.

The approach is motivated by comparisons of existing portfolio selection strategies which have proven to feature components that are beneficial for out-of-sample forecasting performance. Through a 'fair' comparison of popular benchmark approaches under realistic scenarios, I additionally contribute to the financial portfolio modeling literature by showing that relative predictive performances exhibit substantial temporal instabilities over long time

horizons. In particular, I show that in rolling-window type experiments, a general superiority of dynamic over static models cannot be empirically demonstrated, and that the length of the estimation window, which is usually not a focus in econometric papers, should be carefully chosen when estimating a covariance matrix. These insights are used for the development of my proposed model and forecasting approach: First, compared to existing work on graphical LASSO for precision matrices in financial applications, the model enables to account for evolutionary patterns in the correlation structure over time. However, the model does not represent a dynamic modeling approach for conditional (inverse) covariance matrices. Rather, I improve upon unconditional covariance estimation by allowing more remote data information to be considered, but without giving too much weight to the remote observations. Instead, in the vein of Bodnar et al. (2021), only strong deviations from previous timestamps are penalized. Therefore, I incorporate the flexibility of dynamic conditional estimates into the robust estimation procedures for unconditional covariances. Second, motivated by the result that relative forecasting performances appear to exhibit substantial time-instabilities, I enable local adaption to the intensity and the type of regularization. As an alternative to shrinkage, I employ a more flexible but equally parsimonious regularization that depends on a small set of penalty parameters. Third, for answering the empirical question what level of sparsity, temporal stability, and conditioning information is optimal for forecasting, I develop a dynamic recalibration scheme for the penalty parameters that selects the best model in terms of the most recent out-of-sample performance with regard to the evaluation criterion of interest, say the unconditional variance. Hence, the proposed model which is equipped with a problem-oriented dynamic calibration, selects the conditionally optimal parameterization and thereby implicitly the degree of time stability and the weighting of the historical data information. Monthly recalibration allows a fast adaptation to current economic conditions.

To solve the optimization problem, i.e., the minimization of the negative penalized log-likelihood, in an efficient manner, I use the Alternating Direction Method of Multipliers (ADMM), which allows for the inclusion of customized constraints on the precision matrices and their dynamic evolution that complement the L_1-norm regularization induced by LASSO. By the choice of the penalty functions, the augmented time-varying graphical LASSO (ATVGL) is flexible to accommodate specific features of the problem, e.g., minimum variance allocation predictions. Specific requirements can be easily incorporated. Particularly, I show how to directly constrain the gross exposure of GMVP weights in the context of the inverse covariance estimation process without imposing potentially economically unrealistic sparsity on the graph.

In an empirical application to daily returns on U.S.-traded stocks, I show that the proposed approach is able to outperform all considered static and dynamic benchmark models in terms of minimum variance and mean-variance optimal portfolio forecasts over an out-of-sample period from 1980 to 2019, both on average over the 40-year period and in individual 5-year sub-periods.

Chapter 2.

Predicting the Global Minimum Variance Portfolio

2.1. Introduction

The global minimum variance portfolio (GMVP) allocates a given budget among n financial assets such that the risk for the rate of expected portfolio return is minimized. In contrast to the classical mean-variance optimal portfolio (Markowitz, 1952), the weights of the GMVP do not depend on the assets' expected returns. These expected returns have a major impact on the mean-variance optimal strategy (Best and Grauer, 1991) but are notoriously hard to predict (Welch and Goyal, 2008). Studies like Jagannathan and Ma (2003) hence advocate the use of the GMVP, which only depends on the covariance matrix of the asset returns.

In the present paper, we develop a dynamic GMVP approach which takes into account the autoregressive conditional heteroscedasticity of asset returns and aims to predict the conditional GMVP weights, i.e., the weights that minimize the conditional variance of the portfolio returns. To that end, we build upon a consistent loss function (Gneiting, 2011) that arises from a representation of the GMVP weights as population coefficients in an auxiliary linear regression problem (Kempf and Memmel, 2006). We provide a detailed conceptual motivation for a dynamic conditional approach based on this GMVP loss function. In particular, we show that the expected GMVP return entering the GMVP loss function as a nuisance parameter is unavoidable, i.e., there can be no loss function that uniquely identifies the GMVP weights on their own. This result further justifies the use of this loss function. Furthermore, we argue that the expected GMVP loss will typically be very similar to the unconditional portfolio variance, which is the key criterion for evaluating GMVP portfolios in practice. At the same time, the use of a loss function (instead of the unconditional variance) enables tailored model development, parameter estimation and forecast evaluation via Diebold-Mariano tests and related tools. Finally, theoretical results by Ferson and Siegel (2001) indicate that in realistic scenarios with pronounced temporal variation of the conditional covariance matrix of the asset returns and significantly lower variation of their conditional expectations, it is to be expected that dynamic conditional approaches outperform standard static strategies, i.e., approaches that do not explicitly

account for conditional heteroscedasticity.

We then use the GMVP loss function to develop time series models for the conditional GMVP weights, relying on recursive least squares (RLS) with forgetting (Ljung and Söderström, 1983) or generalized autoregressive score (GAS; Creal et al., 2013) recursions. Our approach combines a direct dynamic parametrization of the weights with a consistent loss function for estimating the parameters and evaluating the resulting predictions. This unified setup is in contrast to much existing work that uses different loss functions for estimating versus evaluating an econometric model. In combining a 'non-standard' but consistent loss function with GAS specifications we follow Patton et al. (2019) who consider dynamic models for expected shortfall and value-at-risk.

When the number of assets is large a dynamic GMVP model requires a parsimonious parametrization in order to be tractable in practice. Furthermore, the model should be invariant to the choice of baseline asset in the GMVP loss function. While necessary, this choice is arbitrary and hence should not affect the model's predicted GMVP weights. These two requirements motivate our proposed dynamic GMVP models. Our RLS version of the GMVP model is parsimonious by construction, and in the GAS version parsimony is achieved by specifying a sparse updating recursion targeting the equally weighted portfolio. We show that these RLS and GAS specifications are invariant w.r.t. the baseline asset.

We then present empirical out-of-sample results for various portfolio sizes n and lengths of estimation windows T. In doing so, we cover various realistic setups regarding the ratio n/T which represents the degree of estimation uncertainty for the portfolio weights. Our dynamic GMVP approach, especially the RLS version, performs very well compared to a wide range of benchmarks from the literature, including the dynamic conditional correlation (DCC) model with nonlinear shrinkage estimation of the correlation-targeting matrix (Engle et al., 2019). Furthermore, our proposed loss function is empirically very similar to the empirical (unconditional) portfolio variance, i.e., the popular GMVP performance measure in practice. Hence the benefits of using the loss function (in particular, the ability to use tailored parameter estimation techniques and Diebold-Mariano type forecast evaluation) pose little or no costs in terms of practical relevance.

Existing strategies for predicting GMVP weights can be classified into three types. The first type considers dynamic models for the conditional covariance matrix of the asset returns and constructs a plug-in prediction of the conditional GMVP weights from a forecast of the covariance matrix (see, e.g., Engle and Kelly 2012, Clements et al. 2015, Engle et al. 2019). This approach is designed to account for the time variation in the assets' covariance structure. The covariance models are typically estimated using (quasi) maximum likelihood (ML) or related techniques. Clearly, if the covariance model is correctly specified, ML asymptotically identifies the correct model, which results in optimal forecasts of the GMVP weights. However, this approach can be problematic if the model is misspecified, such that not all properties of the data-generating process (DGP) are correctly captured. Then the

estimator resorts to minimizing a measure of discrepancy between the covariance model and the data (such as the Kullback-Leibler divergence in the ML case). However, such measures are not directly related to the economic problem at hand so that there is the risk that the fitted mis-specified model fails to capture the economically critically important properties of the DGP while at the same time capturing irrelevant ones (Elliott et al., 2016).

The second type of strategies uses a static approach for the GMVP based on either the sample covariance matrix of the asset returns or a shrinkage version thereof (see, e.g., Ledoit and Wolf, 2003, 2004; DeMiguel et al., 2009; Candelon et al., 2012). As we detail further below, such a static approach is implicitly based on the Kempf and Memmel (2006) loss function for the GMVP weights, and thus relies on an estimation principle which under mis-specification appears to be advantageous compared to likelihood-based estimation that is typically used for the plug-in strategies mentioned above. Of course, static approaches do not take into account the conditional heteroscedasticity of asset returns. Therefore, they do not explicitly use conditioning information about future returns, which, as we will discuss below, could be useful for portfolio allocation.

A third group of studies models the portfolio weights as functions of potentially relevant state variables such as firm characteristics; see Brandt (2010, Section 4) for a review. The optimal relation between the portfolio weights and the state variables are found by optimizing a pre-specified utility function. However, in this approach, dynamics in the portfolio weights are accounted for only indirectly via time variation in the state variables.

Our proposed approach combines what we consider the strengths of those three groups of studies: The time series dynamics featured by the first group, the loss function perspective taken by the second and third group, and the third group's proposal to model the weights directly. In contrast to the third group of studies, we construct dynamic models for the conditional weights using pure autoregressive specifications (with current weights assumed to depend on past weights). This modeling approach allows us to approximate key properties of the conditional GMVP weights, such as their persistence or long-run average.

The remainder of this paper is organized as follows. Section 2.2 introduces the GMVP loss function and provides theoretical results motivating our loss-function based dynamic approach. Section 2.3 introduces RLS and GAS models for forecasting the conditional GMVP weights. Section 2.4 presents empirical results, and Section 2.5 concludes with some discussion. Proofs and additional empirical results are deferred to the Appendix.

2.2. Consistent loss function for the GMVP

2.2.1. Setup

Let $R_t = (R_{1t}, ..., R_{nt})'$ denote a vector of returns on n assets at period t. For ease of exposition, we initially assume that R_t is independent across time with expectation $\mathbb{E}[R_t] = \mu$ and covariance matrix $\mathbb{V}[R_t] = \Sigma$. The vector of weights representing the GMVP for the n

assets is denoted by $\omega^* = (\omega_1^*, \dots, \omega_n^*)'$ and obtains as

$$\omega^* = \frac{\Sigma^{-1} \iota_n}{\iota_n' \Sigma^{-1} \iota_n}, \qquad \iota_n' \omega^* = 1, \tag{2.1}$$

where ι_n is an $n \times 1$ vector of ones. According to Kempf and Memmel (2006, Proposition 1) the GMVP weights ω^* can be represented using the following auxiliary linear regression:

$$Y_t = X_t' \beta + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t | X_t] = 0, \tag{2.2}$$

where $Y_t = R_{nt}$ is the return of an (arbitrarily selected) baseline asset and the vector $X_t = (1, R_{nt} - R_{1t}, \dots, R_{nt} - R_{n-1t})'$ consists of the return differences. The corresponding population regression coefficients are $\beta = (\beta_0, \dots, \beta_{n-1})'$, and are defined by

$$\beta = \arg \min_b \mathbb{E}[L(b, R_t)], \tag{2.3}$$

$$\text{where} \qquad L(b, R_t) = (Y_t - X_t' b)^2. \tag{2.4}$$

The population coefficients for the slopes in the auxiliary regression (2.2) coincide with the true GMVP weights and the intercept with the expected GMVP return, in that

$$\omega_i^* = \begin{cases} \beta_i & i = 1, \dots, n-1 \\ 1 - \iota_{n-1}' \beta_{1:n-1} & i = n \end{cases}, \qquad \mu' \omega^* = \beta_0, \tag{2.5}$$

where $\beta_{1:n-1}$ represents the subvector $(\beta_1, \dots, \beta_{n-1})'$ so that $\beta = (\beta_0, \beta_{1:n-1}')'$. In the terminology of Gneiting (2011), $L(b, R_t)$ as given by (2.4) is a strictly consistent scoring (or loss) function for β, which implies that β is elicitable. (A functional of a probability distribution is called elicitable if there exists a strictly consistent loss function for this functional.)

The intercept β_0 is the only element of β that is not directly related to the GMVP weights. It therefore appears tempting to formulate an alternative loss function that exclusively elicits the GMVP weights $\beta_{1:n-1}$. However, as we state in the following proposition such a loss function does not exist. (For the proof, see Appendix A.1.)

Proposition 1. *The sub-vector $\beta_{1:n-1}$ is not elicitable. That is, there is no loss function whose expected value is uniquely minimized by $\beta_{1:n-1}$.*

Clearly, a loss function that elicits $\beta_{1:n-1}$ without β_0 would be attractive in principle. By stating that such a loss function does not exist, Proposition 1 motivates the use of Kempf and Memmel's loss function as a feasible and (strictly) consistent choice. We also note that the joint elicitability of $\beta = (\beta_0, \beta_{1:n-1}')'$ parallels the results on the joint elicitability of value-at-risk and expected shortfall (Fissler and Ziegel, 2016): Just like the GMVP weights $\beta_{1:n-1}$ are not elicitable without β_0, expected shortfall is not elicitable without value-at-risk.

So far, we have assumed that the returns R_t are temporally independent with a constant

covariance matrix. However, it is well-known that financial returns are typically serially dependent with autoregressive conditional heteroscedasticity. Let their corresponding conditional mean and covariance matrix be denoted by $\mu_t = \mathbb{E}[R_t|\mathcal{F}_{t-1}]$ and $\Sigma_t = \mathbb{V}[R_t|\mathcal{F}_{t-1}]$, respectively, where $\mathcal{F}_{t-1}$ is the information set known at time $t-1$. To account for conditional heteroscedasticity, we can then replace the constant covariance matrix Σ in Equation (2.1) by the conditional one Σ_t. Time variation in Σ_t typically implies time variation in the resulting conditional GMVP weights. In this dynamic context, the time-dependent population coefficients of the GMVP regression (2.2) are defined by

$$\beta_t = \arg \min_b \mathbb{E}_{t-1}[L(b, R_t)],$$

where $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$. Analogously to Equation (2.5), it follows that

$$\omega_{it}^* = \begin{cases} \beta_{it} & i = 1, \ldots, n-1 \\ 1 - \iota_{n-1}' \beta_{1:n-1t} & i = n \end{cases}, \qquad \omega_t^{*\prime}\mu_t = \beta_{0t},$$

where $\omega_t^* = (\omega_{1t}^*, \ldots, \omega_{nt}^*)'$ is the vector of portfolio weights,

$$\omega_t^* = \frac{\Sigma_t^{-1}\iota_n}{\iota_n'\Sigma_t^{-1}\iota_n}, \qquad \iota_n'\omega_t^* = 1, \tag{2.6}$$

that minimize the conditional portfolio variance $\mathbb{V}_{t-1}[\omega_t' R_t] = \omega_t'\Sigma_t\omega_t$, and β_{0t} represents the conditional expectation of the corresponding GMVP return.

We use this dynamic GMVP framework based on the consistent loss function $L(b, R_t)$ to develop predictive models for the conditional GMVP weights ω_t^*, to estimate their parameters and to evaluate their predictive performance. The models consist of parametric functions for the weights $\beta_{1:n-1t} = \omega_{1:n-1t}^*$ and expected GMVP return β_{0t}. These functions are assumed to be measurable w.r.t. the information set $\mathcal{F}_{t-1}$, i.e.,

$$(\beta_{0t}, \beta_{1:n-1t}')' = \beta_t = \beta(Z_{t-1}; \theta), \qquad t = 1, \ldots, T, \tag{2.7}$$

where θ is a parameter vector indexing the model, T is the sample size, and $Z_{t-1} \in \mathcal{F}_{t-1}$. In the terminology of Patton et al. (2019) these GMVP models are semi-parametric as they do not impose a parametric class of conditional distributions on the asset returns. For estimating θ we follow Patton et al. (2019) and use an M-type estimator which minimizes the average GMVP loss, so that

$$\hat{\theta} = \arg \min_\theta \frac{1}{T} \sum_{t=1}^T L\left(\beta\left(Z_{t-1}; \theta\right), R_t\right). \tag{2.8}$$

Given our choice of $L\left(\beta\left(Z_{t-1}; \theta\right), R_t\right)$ defined in Equation (2.4), $\hat{\theta}$ is a nonlinear least squares (NLS) estimator.

2.2.2. The unconditional portfolio variance for the conditional GMVP loss approach

As described above, we use the consistent loss function $L(b, R_t)$ in Equation (4) to develop and evaluate predictive models for the weights w_t^* that minimize the *conditional* portfolio variance. This approach reinterprets the standard static GMVP approach, assuming that there is relevant conditioning information $\mathcal{F}_{t-1}$ about future returns. In practice, however, portfolio managers are typically evaluated on the basis of the *unconditional* portfolio variance. This *unconditional* variance is also commonly used to evaluate prediction rules for the *conditional* GMVP weights (see DeMiguel et al., 2009 or Engle et al., 2019 for empirical examples, and Voev, 2009 for a discussion in the context of realized volatility forecasting). This raises two related questions. (i) How does the GMVP loss in Equation (4) compare to the unconditional portfolio variance? (ii) How does our proposed conditional GMVP approach perform in terms of the unconditional portfolio variance, as compared to the traditional static GMVP approach based on an estimate of the returns' unconditional covariance matrix? We tackle these questions in the remainder of this section.

Expected GMVP loss versus unconditional portfolio variance

Let $\{b_t\}$ be a sequence of vectors with arbitrary period-t portfolio weights $\omega_{1:n-1t} = b_{1:n-1t}$ and intercept parameters b_{0t}. According to the law of total variance, the unconditional portfolio variance can be represented as

$$\mathbb{V}[\omega_t' R_t] = \mathbb{E}\left[\omega_t' \Sigma_t \omega_t\right] + \mathbb{V}[\omega_t' \mu_t]. \tag{2.9}$$

The conditional expectation of the period-t GMVP loss can be written as $\mathbb{E}_{t-1}[L(b_t, R_t)] = \omega_t' \Sigma_t \omega_t + (\omega_t' \mu_t - b_{0t})^2$, which together with Equation (2.9) implies that

$$\mathbb{E}\left[L(b_t, R_t)\right] = \mathbb{V}\left[\omega_t' R_t\right] - \mathbb{V}\left[\omega_t' \mu_t\right] + \mathbb{E}\left[(\omega_t' \mu_t - b_{0t})^2\right]. \tag{2.10}$$

Equation (2.10) states that this evaluation of ω_t based on $\mathbb{E}\left[L(b_t, R_t)\right]$ differs from that using the unconditional portfolio variance $\mathbb{V}[\omega_t' R_t]$. The following special cases are worth mentioning here. First, if the conditional mean of the portfolio return is correctly specified, i.e., if $b_{0t} = \omega_t' \mu_t$, then the expected loss simplifies to $\mathbb{E}[L(b_t, R_t)] = \mathbb{V}[\omega_t' R_t] - \mathbb{V}[\omega_t' \mu_t]$, so that the difference between the two performance measures is reduced to the variation in the conditional mean of the portfolio returns. Second, if b_{0t} is constant over time, with $b_{0t} = \omega_t' \mu_t = \mathbb{E}[\omega_t' R_t] \,\forall t$, the difference between $\mathbb{E}\left[L(b_t, R_t)\right]$ and $\mathbb{V}[\omega_t' R_t]$ collapses and both measures are reduced to $\mathbb{E}[\omega_t' \Sigma_t \omega_t]$. Since this is the case for $b_{0t} = 0, \mu_t = 0 \,\forall t$, we expect that both performance measures will yield similar results if the signal-to-noise ratio in the return process is small with large conditional variances in Σ_t and small conditional expectations μ_t. In practice, this is the typical scenario, especially for stock returns at a comparatively

high frequency (such as for daily or weekly returns). However, an important conceptual and practical advantage of $\mathbb{E}\left[L(b_t, R_t)\right]$ is that it corresponds to a standard loss function that can be consistently estimated by the average loss over time. Therefore, it can directly be used for estimating parameters of prediction models and for comparing forecasts in terms of their average out-of-sample loss. The latter is of central importance for commonly used pairwise (Diebold and Mariano, 1995; Giacomini and Rossi, 2010) and multiple (Hansen et al., 2011) comparisons of predictive ability.

Conditional versus static GMVP approach

We next compare the dynamic conditional GMVP approach to the standard static one in terms of their respective unconditional portfolio variance. While the conditional approach exploits the conditioning information $\mathcal{F}_{t-1}$ about future returns in its portfolio allocation, it targets the 'wrong' objective function (conditional instead of unconditional portfolio variance). In contrast, the static approach aims at the 'correct' objective function, but – by using the returns' unconditional covariance matrix – it does not take explicitly into account the conditioning information. Ideally, one would thus aim for an approach that exploits conditioning information *and* uses the 'correct' objective function. Such an approach is found in Ferson and Siegel (2001, Corollary to Theorem 3) who provide the portfolio allocation rule $\omega_t^{\mathrm{FS}} = \omega(Z_{t-1})$, $Z_{t-1} \in \mathcal{F}_{t-1}$, which minimizes the unconditional portfolio variance in Equation (2.9) *with respect to the information set* $\mathcal{F}_{t-1}$. This rule is given by

$$\omega_t^{\mathrm{FS}} = \frac{\Lambda_t^{-1}\iota_n}{\iota_n'\Lambda_t^{-1}\iota_n} + \frac{\gamma_1}{1-\gamma_2}\left(\Lambda_t^{-1} - \frac{\Lambda_t^{-1}\iota_n\iota_n'\Lambda_t^{-1}}{\iota_n'\Lambda_t^{-1}\iota_n}\right)\mu_t, \tag{2.11}$$

$$\text{with} \quad \gamma_1 = \mathbb{E}\left[\frac{\iota_n'\Lambda_t^{-1}\mu_t}{\iota_n'\Lambda_t^{-1}\iota_n}\right], \qquad \gamma_2 = \mathbb{E}\left[\mu_t'\left(\Lambda_t^{-1} - \frac{\Lambda_t^{-1}\iota_n\iota_n'\Lambda_t^{-1}}{\iota_n'\Lambda_t^{-1}\iota_n}\right)\mu_t\right],$$

where $\Lambda_t = \Sigma_t + \mu_t\mu_t'$, and the associated unconditional portfolio variance is

$$\sigma_{\mathrm{FS}}^2 = \mathbb{V}[\omega_t'R_t]\Big|_{\omega_t=\omega_t^{\mathrm{FS}}} = \mathbb{E}\left[\frac{1}{\iota_n'(\Sigma_t + \mu_t\mu_t')^{-1}\iota_n}\right] - \frac{\gamma_1^2}{1-\gamma_2}.$$

While this allocation rule is theoretically appealing, it is not readily applicable in practice: First, it is unclear which loss function could be used to develop predictive models for the optimal weights ω_t^{FS} and to identify the model parameters. Second, plug-in predictions according to Equation (2.11) would require predictive models for μ_t and Σ_t as well as reliable estimates for the moments γ_1 and γ_2 involving functions of μ_t and Σ_t. Such estimates can be difficult to obtain, especially in high dimensional settings (e.g., Ledoit and Wolf, 2003). Importantly, however, the Ferson-Siegel (FS) allocation rule and its variance σ_{FS}^2 provide an infeasible gold standard against which we can compare the conditional and static GMVP approaches.

Representing the returns' unconditional covariance matrix as $\Sigma = \mathbb{E}[\Sigma_t] + \mathbb{V}[\mu_t]$, the un-

conditional portfolio variance in Equation (2.9) for the weights of the static GMVP allocation ω^* in Equation (2.1) can be written as

$$\sigma_{\text{S}}^2 = \mathbb{V}[\omega_t' R_t]\Big|_{\omega_t = \omega^*} = \frac{1}{\iota_n'(\mathbb{E}[\Sigma_t] + \mathbb{V}[\mu_t])^{-1}\iota_n},$$

while for the conditional GMVP weights ω_t^* in Equation (2.6) we have

$$\sigma_{\text{C}}^2 = \mathbb{V}[\omega_t' R_t]\Big|_{\omega_t = \omega_t^*} = \mathbb{E}\left[\frac{1}{\iota_n'\Sigma_t^{-1}\iota_n}\right] + \mathbb{V}\left[\frac{\iota_n'\Sigma_t^{-1}\mu_t}{\iota_n'\Sigma_t^{-1}\iota_n}\right].$$

The comparison of the static, conditional, and FS approaches in different scenarios for the conditional moments μ_t and Σ_t shows the following: (i) For the trivial case that there is no conditional information about future returns so that $\mu_t = \mu = \mathbb{E}[R_t]$ and $\Sigma_t = \Sigma = \mathbb{V}[R_t]$, the three approaches coincide and it holds that $\sigma_{\text{FS}}^2 = \sigma_{\text{S}}^2 = \sigma_{\text{C}}^2 = 1/(\iota_n'\Sigma^{-1}\iota_n)$. (ii) Assume that there is time variation in Σ_t and $\mu_t = m_t\iota_n$, where m_t is a scalar so that the conditional return expectation is the same for all assets. Then the conditional expectation of the portfolio returns $\omega_t'\mu_t$ is identical for all admissible weight vectors ω_t (with $\omega_t'\iota_n = 1$). The Sherman-Morrison formula then implies that

$$\sigma_{\text{S}}^2 = \frac{1}{\iota_n'\mathbb{E}[\Sigma_t]^{-1}\iota_n} + \mathbb{V}\left[m_t\right], \qquad \sigma_{\text{C}}^2 = \mathbb{E}\left[\frac{1}{\iota_n'\Sigma_t^{-1}\iota_n}\right] + \mathbb{V}\left[m_t\right],$$

so that $\sigma_{\text{FS}}^2 = \sigma_{\text{C}}^2 < \sigma_{\text{S}}^2$. By Jensen's inequality, the difference between σ_{C}^2 and σ_{S}^2 increases in the variation of Σ_t. This scenario includes the case that $m_t = \bar{m}\ \forall t$, with the further special case that $m_t = 0\ \forall t$. (iii) If there is variation in μ_t but not in Σ_t such that $\Sigma_t = \bar{\Sigma}\ \forall t$, then

$$\sigma_{\text{S}}^2 = \frac{1}{\iota_n'\Sigma^{-1}\iota_n}, \qquad \sigma_{\text{C}}^2 = \frac{\iota_n'\bar{\Sigma}^{-1}}{\iota_n'\bar{\Sigma}^{-1}\iota_n}\Sigma\frac{\bar{\Sigma}^{-1}\iota_n}{\iota_n'\bar{\Sigma}^{-1}\iota_n},$$

with $\sigma_{\text{FS}}^2 < \sigma_{\text{S}}^2 < \sigma_{\text{C}}^2$, and a difference between σ_{C}^2 and σ_{S}^2 which is increasing in $\mathbb{V}[\mu_t] = \Sigma - \bar{\Sigma}$.

These results reveal that the relative merits of the conditional and static GMVP approaches depend on whether time variation in Σ_t or variation in μ_t (either across time or across its elements) is more pronounced. If the variation in Σ_t is more pronounced, as represented by case (ii), the conditional approach performs better and approaches the FS benchmark as the variation of Σ_t increases. If the variation in μ_t is more important, as in case (iii), the static approach is preferable.

We next quantify these effects using a numerical example calibrated to empirical data. Figure 2.1 plots the unconditional portfolio variance of the three GMVP approaches as a function of the degree of variation in μ_t and Σ_t. The portfolio variances are computed from sequences of conditional return expectations and covariance matrices, $\{\mu_t^*(k)\}_t$ and $\{\Sigma_t^*(\ell)\}_t$,
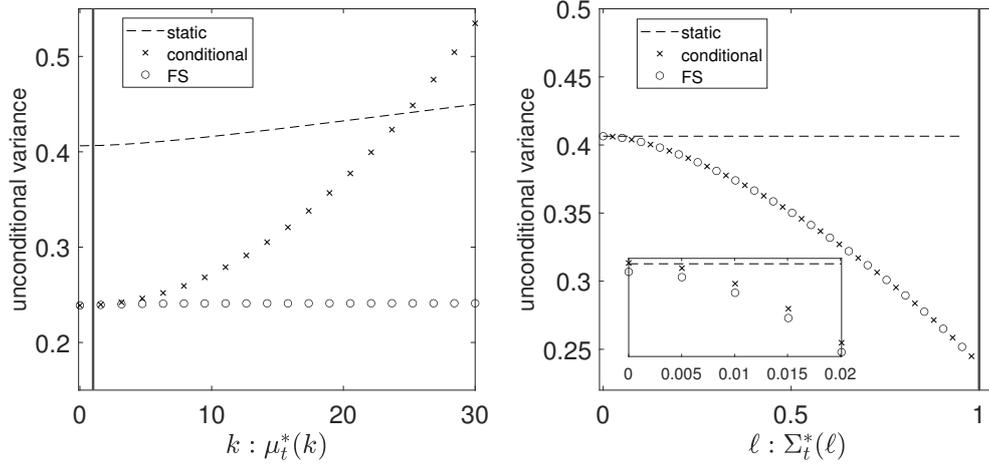
Figure 2.1.: Unconditional portfolio variance for the static, conditional and FS GMVP. Portfolio variances are computed from simulated conditional return expectations $\mu_t(k)$ and covariance matrices $\Sigma_t(\ell)$ defined in Equations (2.12) and (2.13) for different values of k and ℓ. Portfolio variance for $\mu_t(k)$ and $\Sigma_t(\ell)$ with $k = \ell = 1$ are those for μ_t and Σ_t simulated from a fitted Gaussian DCC model with AR(1) recursions for μ_t (marked by the bold vertical line). Simulation sample size for $\{\mu_t(k), \Sigma_t(\ell)\}_{t=1}^{T_s}$ is $T_s = 500,000$ after 50,000 burn-in periods.

which are simulated according to

$$\mu_t^*(k) = \mathbb{E}[\mu_t] + k\delta_t, \quad \mathbb{E}[\delta_t] = 0, \quad k \in [0, 30], \tag{2.12}$$

$$\Sigma_t^*(\ell) = \mathbb{E}[\Sigma_t]^{1/2}[\ell U_t + (1 - \ell)I_n]\mathbb{E}[\Sigma_t]^{1/2'}, \quad \mathbb{E}[U_t] = I_n, \quad \ell \in [0, 1], \tag{2.13}$$

where I_n denotes the $(n \times n)$-dimensional identity matrix and $\mathbb{E}[\Sigma_t]^{1/2}$ is the lower triangular Cholesky factor of $\mathbb{E}[\Sigma_t]$. To simulate $\mu_t^*(k)$ and $\Sigma_t^*(\ell)$ for given values of k and ℓ we use sequences for the conditional moments $\{\mu_t\}$ and $\{\Sigma_t\}$ which are simulated from a fitted Gaussian DCC model (see Section 2.3.4 below) with independent AR(1) recursions for μ_t. The data used are historical returns on $n = 50$ assets from the data set for our empirical application discussed below. From those μ_t's and Σ_t's simulated from the fitted DCC we computed the sample estimates for $\mathbb{E}[\mu_t]$ and $\mathbb{E}[\Sigma_t]$ as well as the corresponding simulated trajectories for $\delta_t = \mathbb{E}[\mu_t] - \mu_t$ and $U_t = \mathbb{E}[\Sigma_t]^{-1/2}\Sigma_t\mathbb{E}[\Sigma_t]^{-1/2'}$ so that $\mu_t^*(k)$ and $\Sigma_t^*(\ell)$ in Equations (2.12) and (2.13) for $k = \ell = 1$ represent the actual empirical (μ_t, Σ_t)-scenario. We then use these δ_t and U_t trajectories to simulate $\mu_t^*(k)$ and $\Sigma_t^*(\ell)$ for other values of k and ℓ. Values $k > 1$ $(k < 1)$ correspond to more (less) temporal and cross-sectional variation in the conditional mean vector, as compared to the empirical scenario, while values $\ell < 1$ correspond to less time variation in the conditional covariance matrix, again as compared to the empirical scenario.

The left panel of Figure 2.1 plots the portfolio variance as a function in k controlling the variation in $\mu_t^*(k)$ for a fixed variation in $\Sigma_t^*(\ell)$ at $\ell = 1$. The figure's right panel plots the

17

portfolio variance as a function in ℓ controlling the variation in $\Sigma_t^*(\ell)$ for a fixed variation in $\mu_t^*(k)$ at $k = 1$. We find that for the empirical benchmark values $\mu_t^*(1)$ and $\Sigma_t^*(1)$, the conditional GMVP improves substantially over the static GMVP with a variance reduction of about 40% (from 0.405 to 0.235) and comes close to the FS lower bound. We also observe that it requires extreme deviations from these empirical benchmark values (i.e., very large values of k or very small values of ℓ) for the static GMVP to outperform the conditional approach.

The lesson we can draw from this comparison is that, in empirically relevant scenarios, our conditional GMVP approach can be expected to perform comparatively well in terms of the unconditional portfolio variance, even though it targets the conditional portfolio variance. It should be noted, however, that our comparison has ignored model misspecifications and parameter estimation errors. While both issues are known to matter in practical forecasting applications, they are hard to analyze via simulation. We therefore limit ourselves to an implicit analysis of these issues in the empirical analysis discussed in Section 2.4 below.

2.3. Dynamic GMVP models

In this section we describe our proposed semi-parametric GMVP models. The first model is based on RLS with forgetting while the second one belongs to the class of GAS models. In both models, the M-estimator from Equation (2.8), together with the GMVP loss function, allows for convenient parameter estimation. For the statistical properties of the M-estimator in semi-parametric models like ours, see Patton et al. (2019).

2.3.1. Recursive least squares with forgetting

Classical exponential forgetting

RLS with forgetting is a popular approach used to track the parameters in linear regression models when they are time-varying (Ljung and Söderström, 1983; Young, 2011). The RLS estimates with classical exponential forgetting (RLS-EF) for the sequence of parameters $\{\beta_t\}$ in the auxiliary regression (2.2) obtain from the recursion

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \Omega_t^{-1} X_t (Y_t - X_t' \hat{\beta}_{t-1}), \tag{2.14}$$

$$\Omega_t = X_t X_t' + \lambda \Omega_{t-1}, \tag{2.15}$$

so that $\hat{\beta}_{1:n-1t}$ and $\hat{\beta}_{0t}$ are the estimates for the period-t GMVP weights and mean return. Ω_t is called the information matrix and $\lambda \in (0, 1]$ is the forgetting factor which operates as an exponential weight decreasing for more remote observations: Observations τ periods in the past have weight λ^τ in the β_t estimate. λ is typically set to a value slightly less than one (Raftery et al., 2010). For the special case that $\lambda = 1$ with $\Omega_t = \sum_{\tau=1}^t X_\tau X_\tau'$, the RLS

recursion in Equations (2.14) and (2.15) becomes

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \ [\sum_{\tau=1}^{t} X_\tau X_\tau']^{-1} X_t (Y_t - X_t' \hat{\beta}_{t-1}). \tag{2.16}$$

This is the standard formula for updating the ordinary LS estimate $\hat{\beta}_{t-1}$ for the parameters in regression (2.2) computed for the observations $\{Y_\tau, X_\tau\}_{\tau=1}^{t-1}$, when a new pair of observation (Y_t, X_t) is added to the sample and the parameters are assumed to be time-invariant (Harvey, 1993, Section 4.5). Thus, $\hat{\beta}_{1:n-1t}$ in Equation (2.16) corresponds to the estimates for the portfolio weights of the standard static GMVP approach obtained by replacing the return covariance Σ in the GMVP formula (2.1) with the sample covariance matrix obtained up to period t. The computation of $\hat{\beta}_t$ given $\hat{\beta}_{t-1}$ according to Equation (2.16) requires that $t \geq n + 1$.

Another interpretation of RLS with forgetting emerges from the Kalman filter with forgetting for a linear Gaussian state-space model (Kulhavỳ and Zarrop, 1993; Raftery et al., 2010): When taking regression (2.2) as a measurement equation for the time-varying parameters β_t with Gaussian measurement errors ε_t and assuming that β_t follows a Gaussian random walk, then $\hat{\beta}_t$ as obtained from the RLS-EF recursion in Equations (2.14) and (2.15) is equivalent to the expectation of β_t under its filtering distribution resulting from the Kalman filter with forgetting. In this context, the RLS information matrix Ω_t plays the role of the precision matrix of the filtering distribution (Kulhavỳ and Zarrop, 1993).

Our proposed dynamic semi-parametric GMVP model based on this RLS-EF approach consists of the mapping $\beta_{t+1} = \beta(Z_t; \theta) = \hat{\beta}_t$, such that

$$\beta_{t+1} = \beta_t + \Omega_t^{-1} X_t (Y_t - X_t' \beta_t), \qquad t = 1, \dots, T, \tag{2.17}$$

with Ω_t as given by Equation (2.15) and initial conditions (β_1, Ω_0). Here we treat the forgetting factor λ directing Ω_t as an unknown parameter in θ that is to be estimated from the data. Note that if a positive definite matrix is chosen for Ω_0, then the information-update equation (2.15) yields an invertible Ω_t ($t = 1, 2, \dots$) as needed for the uniqueness of the RLS-EF recursion (2.17). Thus, our proposed dynamic RLS-EF model, in contrast to the standard static GMVP approach according to Equation (2.16), can also be used for a sample size $T < n + 1$.

In order to capture the non-trivial dynamic behavior of the conditional GMVP variables in β_t that is to be expected for an autoregressive conditional return covariance matrix Σ_t, the proposed GMVP model employs according to Equation (2.17) an autoregressive update mechanism for β_t. The vector $X_t (Y_t - X_t' \beta_{t-1})$ that drives the updates is proportional to the score of the period-t GMVP loss $\partial L(\beta_t, R_t) / \partial \beta_t$ in Equation (2.4). This vector thus measures the steepest descent direction to improve the local forecast performance in terms of the GMVP loss for the current value of GMVP variables β_t. This direction vector is scaled by the inverse of Ω_t that measures the relative amount of locally available information in

period t with regard to the respective direction in the parameter space. The greater the amount of information already available in one direction, the smaller the corresponding update.

In our empirical applications below, the GMVP weights have very persistent time series behavior to the effect that the estimates for λ are close to one. This implies that the selection of the initial conditions (β_1, Ω_0) in the predictive equation (2.17) can become critical for the out-of-sample forecast performance if the length of the estimation window is small. The importance of the initial conditions can be seen in the following representation of β_{t+1} according to Equation (2.17) as a function of (β_1, Ω_0) (Ljung and Söderström, 1983, p. 21):

$$\beta_{t+1} = \left(\lambda^t \Omega_0 + \sum_{\tau=1}^{t} \lambda^{t-s} X_\tau X_\tau'\right)^{-1} \left(\lambda^t \Omega_0 \beta_1 + \sum_{\tau=1}^{t} \lambda^{t-s} X_\tau Y_\tau\right). \tag{2.18}$$

It shows that β_{t+1} is all the more shrunk towards the prior β_1 the closer λ is to one, the larger Ω_0 and the smaller t. Thus, it is advisable to select (β_1, Ω_0) based on reasonable a-priori assumptions that at the same time guarantee scalability for high-dimensional applications. Moreover, the selected (β_1, Ω_0) should ensure that the GMVP model is invariant w.r.t. the choice of the baseline asset in the GMVP regression (2.2) - a qualification discussed in Section 2.3.3 below. (This excludes, for example, the use of a (scaled) identity matrix for Ω_0, which is a common initialization of RLS algorithms in practice, see Ljung and Söderström, 1983, p. 20).

For such a selection of (β_1, Ω_0) we rely upon the prior assumptions, that the return vector R_t has zero mean and a covariance matrix with equal pairwise correlation coefficients denoted by ρ_R and identical variances σ_R^2. Under these assumptions, the GMVP corresponds to the equally weighted portfolio with weights given by $1/n$, and the mean of its return is equal to zero. Furthermore, the return differences in X_t have, like the returns, equal correlations and identical variances with a covariation matrix given by

$$\mathbb{E}_0[X_t X_t'] = \begin{pmatrix} 1 & 0_{n-1}' \\ 0_{n-1} & \sigma_R^2(1-\rho_R)C \end{pmatrix}, \quad C = I_{n-1} + \iota_{n-1}\iota_{n-1}', \tag{2.19}$$

where 0_{n-1} denotes the $n-1$ dimensional Null vector. Based on these prior assumptions we define our initial conditions through:

$$\beta_1 = (0, \iota_{n-1}'/n)', \quad \Omega_0 = \gamma \begin{pmatrix} 1 & 0_{n-1}' \\ 0_{n-1} & \hat{\sigma}_R^2(1-\hat{\rho}_R)C \end{pmatrix}, \quad \gamma > 0,$$

where $\hat{\rho}_R$ and $\hat{\sigma}_R^2$ are the sample estimates for ρ_R and σ_R^2 based on the data in the estimation period. (For details on estimating ρ_R and σ_R^2, see e.g., Engle and Kelly, 2012 and De Nard, 2020.) The coefficient γ controls the degree of shrinkage on β_{t+1} towards the portfolio weights of the equally weighted portfolio defining β_1 (see Equation 2.18). We treat this coefficient γ as an additional parameter to be estimated from the data. So the parameter vector of the

RLS-EF model to be estimated is $\theta = (\lambda, \gamma)$. The restriction $\gamma > 0$ (together with $\hat{\sigma}_R^2 > 0$ and $\hat{\rho}_R \in (-1, 1)$) guarantees that Ω_0 is positive definite, which ensures the invertibility of Ω_t ($t = 1, 2, ...$) in the RLS-EF equation (2.17).

Our selection of (β_1, Ω_0) is in the spirit of the shrinkage approach of De Nard (2020) which aims to reduce the errors in estimating high-dimensional covariance matrices of asset returns. For this purpose, this approach shrinks the sample covariance matrix to a target that consists of equal variances and covariances. Moreover, the resulting initial value β_1 consisting of the weights of the equally weighted portfolio represents a common benchmark portfolio used by shrinkage approaches for GMVP weights (DeMiguel et al., 2009; Candelon et al., 2012; Frey and Pohlmeier, 2016). Also, since our selection of (β_1, Ω_0) is derived from assumptions for the returns R_t (and not for the return differences in X_t relative to the baseline asset), it automatically satisfies the conditions for a dynamic GMVP model that is invariant w.r.t. the choice of the baseline asset (see Section 2.3.3 below and Appendix A.2).

Regularized exponential forgetting

A critical condition for RLS to perform well in tracking time-varying parameters is the existence of lower and upper bounds on the information matrix Ω_t in the parameter update equation (2.17) (Kulhavỳ and Zarrop, 1993). When the smallest eigenvalues of Ω_t get too close to zero, the parameter tracking in the least excited directions of the n-dimensional space of Ω_t becomes very sensitive to noise, and when the largest eigenvalues rise too much, the tracking ability in the most excited directions is lost. In order to achieve a balance between the tracking ability and the robustness against noise when predicting the GMVP weights, it is therefore necessary that the forgetting factor λ in the RLS-EF information update equation (2.15) is neither too small nor too big. However, in our application of the RLS-EF below, we find that quite often the λ estimate is very close to its upper bound of one (see Appendix A.3). Although this avoids the tracking in the poorly excited directions from becoming highly susceptible to noise, it can eventually lead to the loss of the tracking capability in the strongly excited directions.

In order to improve the balance between the robustness w.r.t. to noise and the tracking capability we follow Kulhavỳ and Zarrop (1993) and consider as an alternative to the RLS-EF model in Equations (2.15) and (2.17) an RLS with a regularized exponential forgetting (RLS-REF). It adds the prior matrix Ω_0 to the information update scheme for Ω_t of the RLS-EF (2.15), which then reads as follows:

$$\Omega_t = X_t X_t' + \lambda \Omega_{t-1} + (1 - \lambda)\Omega_0. \tag{2.20}$$

The eigenvalues of this regularized Ω_t are bounded from below for all t and for all $\lambda \in (0, 1]$ by the smallest eigenvalue of Ω_0. Thus this regularization adds prior information, which bounds Ω_t from below regardless of the λ value. This gives more leeway when estimating λ to prevent an excessive increase of Ω_t (i.e., loss of tracking capability) in strongly excited

directions. (In Appendix A.3.2 we provide the bounds for the regularized and non-regularized Ω_t. There we also illustrate the stabilizing effect of the regularization on the time series of Ω_t by providing time plots for the eigenvalues of Ω_t that result from the implementation of the RLS-EF and RLS-REF models.) The initial conditions (β_1, Ω_0) that we use for the RLS-REF model in Equations (2.17) and (2.20) are the same as for the RLS-EF model so that Equation (2.20) yields positive definite updates of Ω_t for all t.

2.3.2. Generalized autoregressive score model

The proposed RLS GMVP models have only two parameters (λ, γ) which direct the joint autoregressive updates of the weights and the mean return of the GMVP. This makes them scalable w.r.t. the number of assets, but could be too restrictive. As an alternative we thus consider a GAS approach which provides some more flexibility in capturing the dynamic behavior of the GMVP variables. The GAS modeling approach as introduced by Creal et al. (2013) assumes that the random variable to be modeled has a parametric conditional distribution with parameters that follow an autoregression driven by the scaled score of the log-likelihood. Following Patton et al. (2019), we adapt this parametric GAS approach to our semi-parametric framework by replacing the log-likelihood score with the score of the GMVP loss function (2.4). This results in a GMVP model that can be interpreted as a parametric extension of the RLS models. The particular GAS recursion we consider for the mapping $\beta_{t+1} = \beta(Z_t; \theta)$ is

$$\beta_{t+1} = c + B\beta_t + AH_t^{-1}\nabla_t, \qquad t = 1, \ldots, T, \tag{2.21}$$

where c denotes an $(n \times 1)$ parameter vector and A and B are $(n \times n)$ parameters matrices. The vector ∇_t is the score of the GMVP loss function given by $\nabla_t = \partial L(\beta_t, R_t)/\partial \beta_t = -2X_t(Y_t - X_t'\beta_t)$, and H_t is an $(n \times n)$ matrix scaling the score. Thus, for $c = 0$, $B = A = I_n$, the GAS update equation (2.21) for β_t has the same form as the update equation (2.17) for the RLS models, subject to the specific form of the GAS-scaling matrix H_t discussed below.

In order to keep the model parsimoniously parameterized, but to allow the update intensity for the GMVP mean return to differ from that of the portfolio weights, we assume the following diagonal form for the GAS matrices A and B:

$$B = \mathrm{diag}(b_0, b_1, \ldots, b_1), \quad A = \mathrm{diag}(a_0, a_1, \ldots, a_1), \tag{2.22}$$

where the parameters (a_0, b_0) direct the dynamics of the mean return and (a_1, b_1) the dynamics of the portfolio weights. For the intercept vector c we use the restrictions that the long-run mean of the weights corresponds to that of the equally weighted portfolio and that the mean of the portfolio returns is zero. Given that β_t under the GAS recursion (2.21) is covariance stationary, the long-run mean of β_t is $m = (I_n - B)^{-1}c$ (see Creal et al., 2013). Thus, our targeting restrictions imply that $c = (I_n - B)m$, with $m = (0, \iota_{n-1}'/n)'$. This

targeting restriction on c again follows the work of DeMiguel et al. (2009); Candelon et al. (2012); Frey and Pohlmeier (2016), which we have already used to motivate the selection for the initial conditions of the GMVP weights for the RLS approach in Section (2.3.1).

For the scaling matrix H_t, a common choice in parametric GAS models is to use a measure for the predicted local curvature of the observation density based on $\mathcal{F}_{t-1}$ (Creal et al., 2013). Accordingly, we select H_t to be related to the curvature of the loss function w.r.t. β_t as measured by its predicted Hessian, which is given by $\mathbb{E}_{t-1}[\partial^2 L(\beta_t, R_t)/\partial\beta_t\partial\beta_t'] = 2\mathbb{E}_{t-1}[X_tX_t']$. For $\mathbb{E}_{t-1}[X_tX_t']$ we use the predictions obtained from an exponential weighted moving average (EWMA) so that our specification for H_t is

$$H_t = 2\mathbb{E}_{t-1}[X_tX_t'], \quad \mathbb{E}_{t-1}\left[X_tX_t'\right] = \kappa\mathbb{E}_{t-2}\left[X_{t-1}X_{t-1}'\right] + (1-\kappa)X_{t-1}X_{t-1}', \quad \kappa \in (0,1]. \tag{2.23}$$

We treat the smoothing coefficient κ as an unknown parameter to be estimated jointly with the parameters in the GAS recursion (2.21). This then gives a parameter vector, which consists of $\theta = (a_0, a_1, b_0, b_1, \kappa)$. Our treatment of the initial conditions is essentially the same as for the RLS models. So we use $\beta_1 = (0, \iota_{n-1}'/n)'$, and for $\mathbb{E}_0[X_1X_1']$, the estimate of the equicovariation matrix $\mathbb{E}_0[X_tX_t']$ in Equation (2.19) obtained by replacing the parameters ρ_R and σ_R^2 by their sample estimates. This ensures that H_t ($t = 1, 2, ...$) in the GAS equation (2.21) is invertible. Accordingly, the GAS version of the GMVP model, just like the RLS versions, can be applied even if $T < n + 1$. Finally note that the GAS scaling matrix H_t in Equation (2.23) is similar to the RLS-EF information matrix Ω_t in Equation (2.15). However, the period-t update of H_t is based on $\mathcal{F}_{t-1}$, in accordance with the standard GAS modeling approach, while Ω_t uses $\mathcal{F}_t$, which implies that the GAS and RLS-EF model are not nested.

2.3.3. Invariance with respect to the selection of the baseline asset

The dynamic GMVP models as described above are specified in terms of the variables (Y_t, X_t) which, in turn, are obtained using asset n as baseline asset (see Equation 2.2). As the choice of the baseline asset is arbitrary, it is desirable for a GMVP model to be invariant w.r.t. this choice. That is, if another asset, say asset k instead of asset n, is chosen as the baseline asset, so that the variables in the GMVP regression (2.2) equal $\tilde{Y}_t = R_{kt}$ and $\tilde{X}_t = (1, R_{kt} - R_{1t}, \ldots, R_{kt} - R_{nt})'$, then the estimated model parameters for both baseline assets n and k should be compatible with each other and lead to the same predictions for the GMVP weights. In Appendix A.2 we show that the proposed RLS and GAS GMVP models with their respective initial conditions exhibit this invariance property.

2.3.4. Ability to track the GMVP weights

In order to analyze the ability of the proposed GMVP models to track the 'true' conditional GMVP weights we conduct a Monte Carlo experiment. The true GMVP weights ω_t^* are generated by simulating a Gaussian DCC model (Engle, 2002). It is given by

$$R_t|\mathcal{F}_{t-1} \sim N(0, \Sigma_t), \quad \Sigma_t = D_t^{1/2} C_t D_t^{1/2}, \tag{2.24}$$

where C_t is the conditional correlation matrix and $D_t = \mathrm{diag}(h_{1t}, \ldots, h_{nt})$ is a diagonal matrix with the conditional return variances, each of which follows a univariate GARCH(1,1) process. The return-specific GARCH(1,1) process is $h_{it} = \varphi_{0i} + \varphi_{1i} R_{it-1}^2 + \varphi_{2i} h_{it-1}$. The correlation matrix C_t is given by $C_t = (Q_t^*)^{-1/2} Q_t (Q_t^*)^{-1/2}$, with $Q_t = (1 - \alpha - \beta)S + \alpha e_{t-1} e_{t-1}' + \beta Q_{t-1}$, where $e_t = D_t^{-1/2} R_t$ is the vector of standardized returns. The diagonal matrix Q_t^* is composed of the diagonal elements of Q_t, and S is the unconditional covariance matrix of e_t. The simulated conditional DCC covariance matrices Σ_t are transformed into the true GMVP weights ω_t^* according to Equation (2.6).

We simulate $M = 500$ multivariate series of n returns R_t each of length T. To each of them, we fit the GMVP models so that we obtain M sequences of in-sample estimates for the vector of GMVP weights. The parameters of the DCC model used as DGP are selected as in the experiment of Engle et al. (2019): The correlation parameters are set to $\alpha = 0.05$ and $\beta = 0.93$, the GARCH(1,1) parameters to $\varphi_{1i} = 0.05$ and $\varphi_{2i} = 0.9$ for all n return variates, and the unconditional covariance matrix S is set equal to its sample estimate which obtains for the n most liquid stocks of our data set using 10 years of daily data from 2007 though 2016 (see Section 2.4.2). We consider three portfolio sizes $n \in \{50, 200, 500\}$ and two lengths of the estimation window $T \in \{250, 1250\}$ which corresponds approximately to one and five years of daily return data. The resulting concentration ratio n/T, which captures the degree of estimation uncertainty, varies from 0.04 to 2. For evaluating the accuracy of the in-sample estimates of the conditional GMVP weights we use the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the estimates relative to the true values.

In Table 2.1, we report the MAE and RMSE results for our GMVP models. For comparison, it also provides the results for GMVP-weight estimates that result from fitting a DCC model to the simulated data and plugging the resulting estimates for the covariance matrix Σ_t into the GMVP formula (2.6). To implement correlation targeting in the DCC matrix Q_t for fitting the DCC model, we use the standard approach (DCC-s), which consists of estimating the unconditional covariance matrix S by the sample covariance matrix of the standardized returns e_t (Engle, 2002). But this is only feasible as long as $n/T \leq 1$, since for $n/T > 1$ the sample covariance matrix does not have full rank. As an alternative, we consider the approach of Engle et al. (2019), who propose to use for the target matrix S the nonlinear shrinkage (nl) estimator of Ledoit and Wolf (2012, 2015). The resulting DCC-nl approach is also feasible for $n/T > 1$. Both DCC models are estimated by a composite

Table 2.1.: Simulation results

MAE	$T = 250$			$T = 1250$		
n	50	200	500	50	200	500
RLS-EF	2.420	4.712	6.965	2.701	5.018	6.744
RLS-REF	2.277	4.662	6.933^+	2.277^+	4.390^+	6.445^+
GAS	$\mathbf{2.252}^+$	$\mathbf{4.634}^+$	6.946	2.415	4.556	6.530
DCC-s	3.188	12.462	–	1.865	5.185	8.812
DCC-nl	2.565	4.684	**6.589**	**1.760**	**4.352**	**6.093**

RMSE	$T = 250$			$T = 1250$		
n	50	200	500	50	200	500
RLS-EF	0.450	0.457	0.445	0.502	0.474	0.418
RLS-REF	$\mathbf{0.424}^+$	$\mathbf{0.454}^+$	$\mathbf{0.416}^+$	0.419^+	$\mathbf{0.416}^+$	0.407^+
GAS	0.426	0.458	0.447	0.450	0.437	0.416
DCC-s	0.608	1.187	–	0.354	0.499	0.537
DCC-nl	0.487	0.455	0.421	**0.334**	0.418	**0.379**

Note: Mean absolute error (MAE) and root mean squared error (RMSE) of the estimated GMVP weights for $M = 500$ simulations. Bold figures indicate the smallest error across all models and figures marked by $^+$ indicate smallest error across models based on the GMVP loss. The MAE is computed as $\sum_{s=1}^M [\sum_{t=1}^T d_{ts}/T]/M$ with $d_{ts} = \sum_{i=1}^n |\omega_{its}^* - \hat{\omega}_{its}|$, where $\hat{\omega}_{its}$ is the estimate for the weight ω_{its}^* of asset i in period t for simulation run s. The RMSE is computed analogously based on $d_{ts} = [\sum_{i=1}^n (\omega_{its}^* - \hat{\omega}_{its})^2]^{1/2}$.

likelihood approach as recommended by Engle et al. (2019).

The results in Table 2.1 show that among the proposed GMVP models, the RLS-REF tends to perform best in tracking the GMVP weights. It dominates the RLS-EF across all n/T scenarios and both performance measures. This suggests that the regularization of the information matrix Ω_t improves the balance between the sensitivity to changes in the GMVP weights and the robustness w.r.t. noise. For the long estimation window, the RLS-REF also dominates the GAS model. Only for the short window and portfolio sizes of $n = 50$ and $n = 200$, the weight estimates of the GAS have lower MAE values. The RLS-REF also performs well relative to the DCC that defines the DGP. Even if the RLS-REF for the long estimation window is typically outperformed by the DCC-nl, it tends to provide more accurate weight estimates than the DCC models for the short window. This can be explained by the fact that the DCC, which has substantially more parameters than the RLS and GAS models, suffers from comparably large parameter estimation uncertainty with small sample sizes, resulting in excess variation of the weight estimates. It is known that the DCC-nl, especially for a large n/T, improves parameter estimation relative to the DCC-s (see Engle et al., 2019), which explains why the DCC-nl performs better than the DCC-s in tracking the weights.

2.4. Empirical application

In this section we apply our proposed GMVP prediction models to historical daily return data, and compare them to a set of benchmark models in an out-of-sample forecasting experiment.

2.4.1. Data and design of the experiment and benchmark models

We use the data set analyzed by Moura et al. (2020) which consists of the daily prices of all NYSE, AMEX and NASDAQ stocks and update it to include price observations until the end of the year 2019. The full sample covers the period from 01/02/2002 to 12/09/2019 for a total of 4516 trading days. In our out-of-sample experiments we focus on one-day-ahead forecasts obtained by re-estimating the model parameters every month on a rolling window scheme, where we follow the convention that 21 consecutive days constitute one month. The out-of-sample period starts on 01/03/2007 and ends on 12/09/2019 which results in a total of 3257 point forecasts. We consider four portfolio sizes, $n \in \{50, 200, 500, 1000\}$, and two lengths of the estimation window, $T \in \{250, 1250\}$ (one and five year). We thus cover eight practically relevant scenarios of n/T ranging from 0.04 to 4.

Following Engle et al. (2019), the n stocks included in a portfolio are re-determined before re-estimating the parameters each (virtual) month. They are selected as follows: First, we identify the stocks that have a complete series of reported returns over the most recent T days and over the next 21 days. Then, we identify all pairs of stocks with a sample correlation larger than 0.95 over the past T days and remove the respective stock with lower trading volume observed at the time of re-estimation. Finally, we select the largest n stocks in terms of market capitalization at the re-estimation period. A time plot of the returns for an equally weighted portfolio including the 1000 largest stocks is provided in Appendix A.3.

As alternatives to our proposed GMVP models we use the following six static and dynamic approaches from the literature: (i) The OLS estimator constructs the period-t prediction of the GMVP weights by replacing the return covariance matrix Σ in the GMVP formula (2.1) with the sample covariance matrix of the returns observed up to period $t-1$. This approach is equivalent to running Kempf and Memmel's (2006) static auxiliary regression using OLS and corresponds to the RLS-EF with a forgetting factor $\lambda = 1$ (see Equation 2.16). (ii) The linear shrinkage (SHR-l) estimator modifies the OLS estimator by estimating Σ in Equation (2.1) via the linear shrinkage approach of Ledoit and Wolf (2004). (iii) The nonlinear shrinkage (SHR-nl) estimator estimates Σ in formula (2.1) via the nonlinear shrinkage procedure of Ledoit and Wolf (2012, 2015). While the linear shrinkage estimator shrinks all sample eigenvalues towards the grand mean of the sample eigenvalues with the same intensity, the nonlinear shrinkage approach uses an individualized intensity for each eigenvalue. (iv) The naïve estimator sets the prediction of the GMVP weights equal to the weights of the equally weighted portfolio. (v) The Gaussian DCC model in Equation (2.24)

with correlation targeting based on the standard sample covariance matrix (DCC-s). (vi) The Gaussian DCC model with correlation targeting based on nonlinear shrinkage estimates (DCC-nl).

2.4.2. Results

Out-of-sample GMVP loss and portfolio variance

In Table 2.2 we report the average out-of-sample GMVP loss of the proposed GMVP models as well as the benchmark models. (The NLS-estimates for the key parameters of the GMVP models are provided in Appendix A.3.) For assessing the statistical significance of differences in the average out-of-sample GMVP loss across models, we use the model confidence set (MCS) approach of Hansen et al. (2011). The MCS is constructed to contain the best-performing models at a given confidence level, which we set equal to 90%. In the bootstrap implementation of the MCS, we use a block bootstrap with block length $\lfloor T_{eval}^{1/3} \rfloor$, where $T_{eval} = 3257$ is the size of the evaluation sample, and a bootstrap sample size of 10000. We use the implementation of the MCS procedure in the Oxford MFE toolbox (`https://www.kevinsheppard.com/code/matlab/mfe-toolbox`). Results of a forecast comparison based on the test for superior predictive ability (Hansen, 2005) are qualitatively very similar and are available in Appendix A.3.

Table 2.2.: Average out-of-sample GMVP loss and variance

			$T = 250$				$T = 1250$		
n		50	200	500	1000	50	200	500	1000
n/T		0.20	0.80	2.00	4.00	0.04	0.16	0.40	0.80
RLS-EF	Avg. loss	0.577	0.421	0.391	0.300	0.559	0.438	0.354	0.297
	Portf. var.	0.574	0.420	0.390	0.297	0.558	0.438	0.354	0.296
RLS-REF	Avg. loss	**0.560**	**0.415**	**0.349**	**0.285**	**0.530**	**0.418**	**0.348**	0.293
	Portf. var.	0.558	0.414	0.348	0.285	0.528	0.418	0.348	0.294
GAS	Avg. loss	0.688	0.548	0.496	0.418	0.587	0.485	0.375	0.333
	Portf. var.	0.693	0.524	0.496	0.399	0.583	0.474	0.370	0.331
DCC-s	Avg. loss	0.694	0.927	–	–	0.635	0.501	0.389	0.395
	Portf. var.	0.691	0.921	–	–	0.631	0.500	0.386	0.392
DCC-nl	Avg. loss	0.638	0.468	0.364	**0.255**	0.628	0.481	**0.335**	**0.260**
	Portf. var.	0.635	0.465	0.362	0.253	0.624	0.478	0.332	0.258
OLS	Avg. loss	0.701	1.483	–	–	0.623	0.495	0.464	0.804
	Portf. var.	0.696	1.471	–	–	0.620	0.493	0.462	0.800
SHR-l	Avg. loss	0.680	0.824	0.506	0.336	0.622	0.493	0.457	0.637
	Portf. var.	0.676	0.817	0.501	0.333	0.618	0.490	0.454	0.633
SHR-nl	Avg. loss	0.630	0.454	0.368	0.321	0.619	0.471	0.382	0.317
	Portf. var.	0.625	0.450	0.365	0.319	0.615	0.468	0.379	0.315
naïve	Avg. loss	1.373	1.538	1.690	1.847	1.366	1.535	1.666	1.851
	Portf. var.	1.363	1.527	1.677	1.834	1.357	1.535	1.654	1.837

Note: Average out-of-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) for portfolio sizes $n \in \{50, 200, 500, 1000\}$. Parameter estimation is based on a sample of length T. Bold numbers indicate smallest average GMVP loss across all models and grey cells indicate that the model belongs to the 90% MCS.

The results for the average out-of-sample GMVP loss in Table 2.2 can be summarized

as follows: For all eight (n, T)-scenarios, the RLS-REF attains the smallest average loss among our proposed GMVP models. The fact that the RLS-REF outperforms RLS-EF is fully in line with the results of the Monte Carlo experiment in Section 2.3.4, which indicate that regularizing the exponential forgetting of RLS enhances the performance in tracking the GMVP weights. Both RLS versions consistently achieve a smaller average loss than the GAS model. The RLS approach thus seems to better capture the characteristics in the dynamics of GMVP weights that are important for its predictions with a similarly small number of parameters than the GAS model (2 in the RLS versus 5 in the GAS). The best performing benchmark models for both estimation-window lengths T are the DCC-nl (for $n = 500$ and $n = 1000$) and the SHR-nl (for $n = 50$ and $n = 200$). We also find that the forecast improvements achieved when moving from the standard DCC-s to DCC-nl and from the standard OLS estimator to SHR-nl consistently increase with n/T. This confirms the results of Engle et al. (2019) and Ledoit and Wolf (2012, 2015) which suggest that nonlinear shrinkage is particularly beneficial in large dimensional applications. The comparison of all competing models reveals that the RLS-REF is the only model that belongs to the 90% MCS for all (n, T)-scenarios. It also has the lowest average loss, except for the scenarios $(n, T) = (500, 1250)$ and $n = 1000$, in which the loss of the DCC-nl is lower. A possible explanation for the lower loss of the DCC-nl compared to the RLS-REF in those large-n scenarios is that the greater flexibility of the highly parameterized DCC is better able to cope with a large heterogeneity in the dynamic behavior of the returns across stocks. Remember that we construct the asset universes for the portfolios based on the assets' market capitalization, so that the additional stocks that are added when the universe is expanded will gradually decrease in size. The fact that this increases the heterogeneity in the behavior of the stock returns is shown by the significant rise in the spread of the estimates for the return-specific GARCH parameters $(\varphi_{0i}, \varphi_{1i}, \varphi_{2i})$ (not reported here), that we find for the DCC when fitted to an increasing number of stocks. Finally, for the short estimation window, the static SHR-nl estimator performs quite well relative to the dynamic approaches. This is to be expected, since the shorter the (rolling) estimation window, the easier it is for static approaches to adapt to local parameter changes.

In addition to the average out-of-sample loss, we report in Table 2.2 the sample variance of the predicted GMVP returns. The comparison of the values for the two performance measures shows that they are very close to each other. The correlation coefficient for their values across the nine competing models and eight (n, T)-scenarios is equal to 0.99995. In addition, the ranking of the models for the two measures is identical in six out of the eight (n, T)-scenarios, and in the two scenarios with different rankings we only see a swap of the ranking for the GAS and the DCC-nl ($n = 200$, $T = 1250$) and for the GAS and the DCC-s model ($n = 50$, $T = 250$). As discussed in Section 2.2.2, this close similarity between the two performance measures is the result of very small signal-to-noise ratios in the process of the stock returns.

Table 2.3.: Average out-of-sample GMVP loss and variance one estimation

	RLS-EF	RLS-REF	GAS	DCC-s	DCC-nl	OLS	SHR-l	SHR-nl	naïve
Avg. loss	**0.564**	0.565	0.675	0.682	0.679	0.995	0.992	0.946	1.363
Portf. var.	0.563	0.564	0.661	0.681	0.678	0.994	0.991	0.944	1.361

Note: Average out-of-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) for portfolio size $n = 50$. For the predictions in the out-of-sample window, model parameters are fixed to their estimated values from the initial estimation window (01/03/2007 – 12/09/2019 with $T = 1250$). Bold numbers indicate smallest average GMVP loss across all models and grey cells indicate that the model belongs to the 90% MCS.

The results in Table 2.2 indicate that the best static GMVP approach, the SHR-nl, performs quite well relative to the dynamic models, especially for the short estimation window. This finding seems to contradict the results of our calibration experiment in Section 2.2.2, which point to clear performance gains of dynamic approaches over static ones. For a portfolio size of $n = 50$, this experiment shows for the empirically relevant benchmark scenario a reduction in the unconditional portfolio variance for the dynamic approach by approximately 40% compared to the static one. In contrast, the results in Table 2.3 for $n = 50$ show that the changes in the portfolio variance for the dynamic approaches compared to the static SHR-nl are in a range from -14% (RLS-REF, $T = 1250$) to +11% (GAS, $T = 250$). This apparent contradiction can be explained by the fact that in our forecast experiment the static approaches sequentially revise their estimates for the covariance matrix Σ based on the updated return history in the rolling estimation window scheme. These estimates hence capture local trends in the conditional covariance matrix Σ_t, while in the calibration experiment, the unconditional covariance matrix Σ used by the static approach is kept fixed by definition. To provide evidence for this interpretation, we modify the out-of-sample experiment by suppressing the revisions of the parameter estimates during the out-of-sample period and fixing the parameters for the static and dynamic models to their estimates from the initial estimation window. The out-of-sample period is the same as before and the length of the initial estimation window is $T = 1250$. As in the calibration experiment, we use a portfolio size of $n = 50$. Table 2.3 reports the results of this modified experiment, showing clear gains for the dynamic models over the SHR-nl and the other static approaches, in line with the calibration experiment in Section 2.2.2. In fact, the reduction in the portfolio variance compared to the SHR-nl achieved by the dynamic models ranges from 28% (DCC-s) to 40% (RLS-EF, RLS-REF) and is thus of a similar magnitude as the reduction found in the calibration experiment.

The out-of-sample GMVP loss and the portfolio variance as considered here are natural performance measures for predictive models of the GMVP weights. In Appendix A.3.4 we also provide the results for the out-of-sample Sharpe ratio. They show that the proposed GMVP models also perform well with regard to this criterion compared to the competing

benchmarks, with the largest Sharpe ratio for five of the eight (n, T)-scenarios considered. However, as emphasized by Engle et al. (2019), this measure is only of secondary importance for the evaluation of prediction models for the GMVP allocation, even if a large Sharp ratio is generally desirable.

Local out-of-sample performance over time

For assessing the models' relative out-of-sample forecasting performance over time, we employ the fluctuation test of predictive ability in unstable environments proposed by Giacomini and Rossi (2010). We use this test to compare the GMVP-weight forecasts of our best performing approach, the RLS-REF, with the best static benchmark model (SHR-nl) and the best dynamic benchmark (DCC-nl). We implemented the fluctuation test using the average out-of-sample GMVP loss within a rolling evaluation window of 328 observations (10% of the full out-of-sample period). As in our main experiment, we obtain the forecasts by re-estimating the model parameters using a rolling window scheme. For each of our (n, T)-combination, Figure 2.2 plots the Giacomini-Rossi (GR) fluctuation test statistic and the 5% asymptotic critical value for the hypothesis that the forecasting performance of the RLS-REF is equal or worse than the benchmark at each point in time, which is rejected if the critical value is crossed at least once. Negative values of the test statistic imply that the RLS-REF forecasts better than the respective competing model. The plots show gains in the out-of-sample performance of RLS-REF compared to the SHR-nl and DCC-nl benchmarks particularly in the 2007-2008 financial crisis, indicating that this important period contributes to the good overall performance of RLS-REF. Beyond the crisis period, the models' relative performance is generally fairly stable, with one notable exception: For $(n, T) = (200, 1250)$, the loss differential of RLS-REF and DCC-nl also exhibits a distinct negative trend in the last years of our out-of-sample period, which indicates that the sparse RLS model can outperform the highly-parameterized DCC also in tranquil economic periods.

2.5. Summary and discussion

In order to develop and evaluate predictive models for the weights of the global minimum variance portfolio (GMVP) which minimize the conditional portfolio variance, we propose to use the Kempf and Memmel (2006) loss function which is defined by an auxiliary linear regression. We provide theoretical arguments to justify the use of this GMVP loss function. The GMVP loss function is strictly consistent for the GMVP weights, and although this consistency only applies in conjunction with the mean of the GMVP returns, we show that there is no loss function that uniquely identifies the GMVP weights on their own. Furthermore, in empirically relevant scenarios the mean of the GMVP loss can be expected to be a very close approximation of the unconditional portfolio variance, which in practice is the predom-
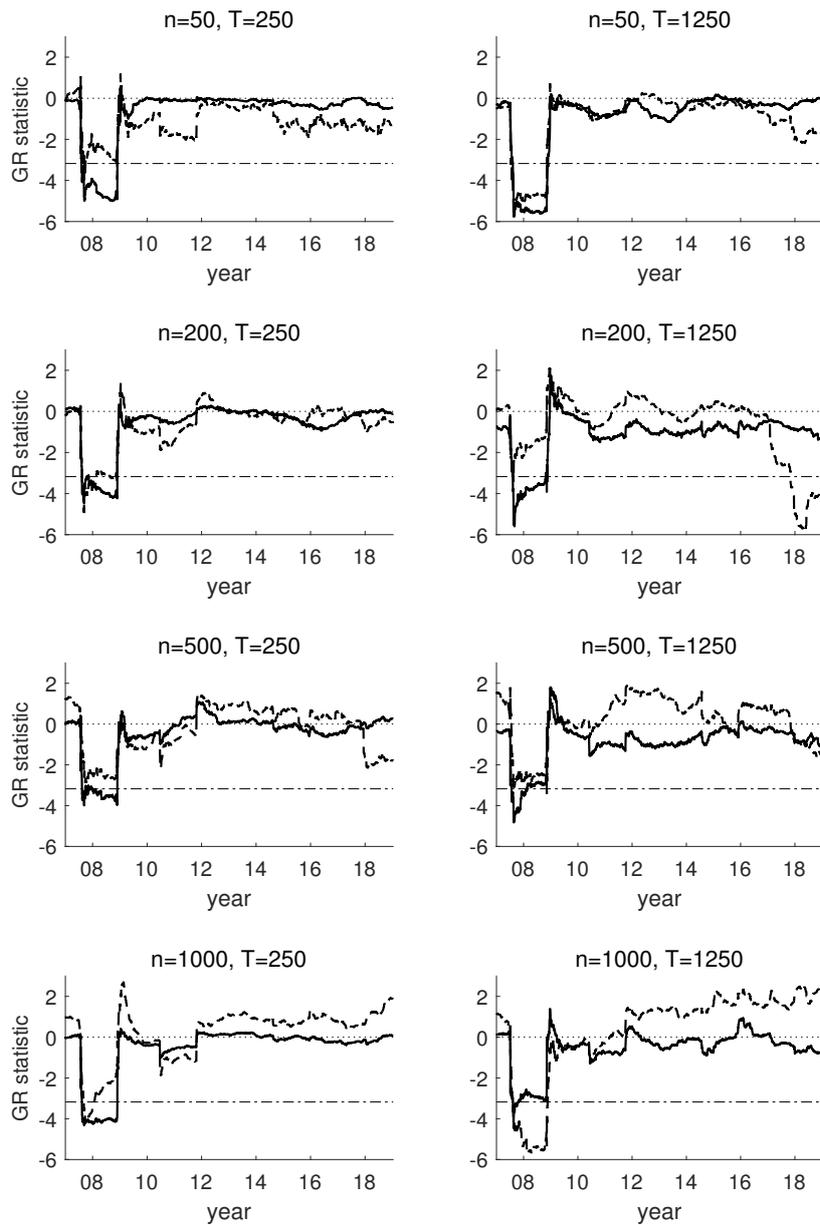
Figure 2.2.: GR fluctuation test statistic for the hypothesis that the RLS-REF is equal or worse than the DCC-nl (bold dashed line) and the SHR-nl (bold solid line). Dashed-dotted line marks the 5% critical value and dotted line the origin.

inant measure for evaluating portfolios targeting the GMVP. We then use the GMVP loss to develop sparsely parameterized dynamic models for the conditional portfolio weights that are easily scalable w.r.t. the size of the portfolio. The proposed models are semi-parametric and based on recursive least squares (RLS) with standard exponential forgetting (EF) or regularized exponential forgetting (REF) as well as generalized autoregressive score (GAS) recursions. The parameters of those models can be easily estimated by standard nonlinear least squares. Using daily stocks prices, we find that the RLS GMVP model with REF in

particular performs very well compared to a wide range of relevant benchmark approaches from the literature and across different portfolio sizes ranging from 50 to 1000 stocks.

There are several avenues for future research. One would be to explore RLS algorithms with alternative forgetting schemes that can improve the ability to track the GMVP weights in scenarios with a very large heterogeneity in the dynamic behavior of asset returns, such as in our applications for the large portfolios. In these cases, the degree of excitation in the different directions of the parameter space of the GMVP regression is typically very heterogeneous (see Appendix A.3.2) so that it could be useful to endow the RLS with a selective directional forgetting (DF) scheme. Such RLS-DF algorithms, which admit a decrease of accumulated information in the information matrix only in the currently excited direction, are proposed by Bittanti et al. (1990) and Kulhavỳ and Zarrop (1993), and further developed, e.g., in Goel and Bernstein (2018) and Shin and Lee (2020). In a preliminary pilot application of RLS with one of the simplest DF schemes we have indeed found improvements relative to the RLS-EF and RLS-REF for our portfolio with 1000 stocks, which confirms our intuition.

Another topic for future research is the treatment of constraints that portfolio managers often have to consider in practice, such as restrictions on the gross exposure, turnover, maximum (absolute) position or factor exposure. All these constraints can be fairly easily incorporated in plug-in GMVP approaches based on predictions of the returns' conditional covariance matrix, namely by solving standard convex optimization problems. To include such constraints in the proposed GMVP models the autoregressive update mechanism for the portfolio weights in the dynamic GMVP regression would need to be modified so that it meets the constraints period by period. An approach which accounts for gross-exposure constraints within the static GMVP regression for a fixed (unconditional) covariance matrix of the asset returns is found in Fan et al. (2012). It exploits the fact that such constraints can be represented as an inequality restriction on the 1-norm of the vector with the portfolio weights. With this 1-norm the GMVP regression can then be formulated as a LASSO-regression problem (least-absolute shrinkage and selection operator) with a corresponding LASSO-solution path for the range of gross-exposure ratios. So one possibility to adapt our dynamic GMVP approach to account for gross-exposure constraints is to combine the GMVP regression with an RLS algorithm with 1-norm constraints. Such regularized RLS algorithms are developed, e.g., by Angelosante et al. (2010) and Nascimento and Zakharov (2016). However, it is unclear how other practical restrictions can be taken into account in addition to a gross-exposure constraint. This is a limitation of our proposed GMVP approach compared to plug-in GMVP strategies.

Chapter 3.

A Time-Varying Parameter Model with Bayesian Shrinkage for Global Minimum Variance Portfolio Prediction

3.1. Introduction

Modern asset allocation considers hundreds of financial assets, and its fundamental input is an estimate of the covariances of returns. However, when the number of cross sectional units n is large relative to the length of the time series T, the sample covariance matrix is subject to significant estimation error. Thus, estimating optimal portfolio weights becomes challenging, both numerically and statistically, when the number of assets considered is large.

We propose a dynamic sparse model directly for the weights of the Global Minimum Variance Portfolio (GMVP), which is particularly suited for situations where the length of the time series of returns is not much larger than the number of assets considered. Operating solely at the quantity of interest for portfolio allocation, we are able to sidestep the direct estimation of large and often ill-conditioned covariance matrices. Exploiting that the GMVP weights can be obtained as the population coefficients of a linear regression allows us to deal with estimation noise and regularization in a straightforward way. We set up a linear state space model with time-varying parameters and stochastic volatility, which addresses both the time variation in the assets' conditional covariance structure and the heteroscedasticity in the market. Bayesian inference techniques with LASSO (least-absolute shrinkage and selection operator) type priors provide data driven shrinkage mitigating the problems of overfitting, large asset exposure, and high turnover.

Conceptually, employing a GMVP strategy is only optimal for an infinitely risk averse investor or under the assumption that the expected returns are equal across all assets. However, several studies present evidence that the GMVP often performs better out-of-sample than other mean-variance portfolios even when performance measures that depend on average returns are considered (see, for example, DeMiguel et al., 2009). Since expected returns have a major impact on the mean-variance optimal strategy (Best and Grauer, 1991), but are notoriously hard to predict (Welch and Goyal, 2008), estimation error is so large that

not much is lost in ignoring differences in expected returns (Jagannathan and Ma, 2003). In addition, a recomposition of the portfolio based on updated estimates of the (conditional) correlation structure typically takes place at regular intervals. Since any rebalancing in practice involves turnover costs, 'stable' portfolios are to be preferred (see, e.g., Li, 2015 for a discussion) so that a high sensitivity of portfolio weights to small changes in expected values also for this reason is unfavourable.

However, when the number of assets is large relative to the number of observations, also estimates of the second moments of asset returns are subject to considerable estimation error, affecting the GMVP weights (Basak et al., 2009). One way to alleviate estimation noise is to impose sparsity restrictions like diagonality (see, e.g., Kirby and Ostdiek, 2012) or a factor structure on the covariance, but these ad hoc restrictions are often false and cannot be revised by the data. In contrast to that, shrinkage of the unconditional covariance or correlation matrices, as proposed in Ledoit and Wolf (2003, 2004, 2012, 2015), does not force sparsity; nevertheless, it is able to reduce estimation noise very effectively. An easy way to incorporate shrinkage procedures in dynamic modeling approaches has been proposed by Engle et al. (2019), who use shrinkage estimates of the unconditional correlation matrix as the targeting matrix in the dynamic conditional correlation (DCC) model of Engle (2002). A third approach, building on the argument that the portfolio weights are a linear combination of the entries of the inverse covariance rather than the covariance matrix itself, is to establish a sparse model for the precision of the data. A prominent example is the approach of Callot et al. (2019), who combine LASSO estimation with a nodewise regression representation for the elements of the precision matrix.

Although the previously discussed approaches are developed to deal with the difficulties of estimating covariances between a large number of cross sectional units, these regularizations do not operate directly on the quantity of interest for portfolio selection. Regularizing portfolio weights is a more direct way to avoid unreasonably high exposures to certain assets, or a high degree of short-selling. In addition to linear shrinkage in which the resulting weights estimates obtain as linear combination of the optimal portfolio rules with some reference (DeMiguel et al., 2013; Tu and Zhou, 2011), and exponential smoothing (Golosnoy et al., 2019), a popular regularization technique is to impose norm constraints on the portfolio weights (see, e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009), typically L_1 or L_2 penalties or combinations thereof (Li, 2015; Yen, 2016). Both of them lead in the limit to the equally weighted portfolio, but the path is different: While under L_1, or LASSO-type, penalty, sparse portfolios are encouraged and the amount of short selling is particularly penalized, L_2 or ridge-type regularization leads to more uniform allocations punishing large exposures (for a discussion see DeMiguel et al., 2009; Ait-Sahalia and Xiu, 2017). Frey and Pohlmeier (2016) implement norm-constraints in a Bayesian setting with ridge and LASSO type priors that shrink weights in direction of a reference portfolio like the equally weighted portfolio. Imposing priors in a particular way corresponds to a probabilistic shrinkage of the

parameter space such that no hard restrictions have to be made a-priori.

In the present paper, we extend the Bayesian shrinkage approach for the portfolio weights to a dynamic context. This means not only that we allow for dynamics in the GMVP weights to reflect instabilities in the conditional correlation structure of the return series, but also that we regulate them via the prior specifications, thus supporting stable portfolio allocation more than other dynamic approaches. Building on an reference augmented linear regression model in which the regression coefficients correspond to the weights deviations of the GMVP relative to some reference portfolio (Kempf and Memmel, 2006; Frey and Pohlmeier, 2016), we set up a time-varying parameter model with stochastic volatility that captures also the conditional heteroscedasticity in the market. An advantage of this approach is that the problem scales linearly such that it is easily applicable to large dimensions and performs well when n/T is large. The resulting linear Gaussian state-space model is easily tractable via standard Bayesian MCMC procedures. Bayesian shrinkage type priors for the dynamics and the overall level of the portfolio weights fulfill two tasks at the same time: First, they reduce estimation noise and thereby allow to cope also with settings under high concentration ratios. Second they regularize the allocations; in particular, when using the time-invariant $1/n$ allocation rule benchmark, they explicitly reduce short selling and exposure as well as the volatility of the weights leading to a smoother evolution of the weights associated with moderate turnover.

The performance of our approach is demonstrated by simulation exercises based on DCC dynamics for the covariance of the assets. Remarkably, the dynamic GMVP approach delivers less risky portfolios and better weight forecasts than the true model when large concentration ratios are considered. Moreover, we conduct an extensive empirical analysis based on a data sets with up to 1600 daily asset returns, which confirms the superior performance of the proposed approach. Notably, although our model is designed to predict the weights of the GMVP, it also delivers satisfying Sharpe ratios.

The remainder of this paper is organized as follows: Section 2 introduces the representation of GMVP weights as time-varying parameters in a linear regression model. In Section 3 we present our proposed reference augmented specification for the GMVP regression and its formulation as a linear Gaussian state space model with time-varying parameters and stochastic volatility. Section 4 describes the Bayesian shrinkage specifications, posterior analysis and forecasting. In Sections 5 and 6 we present the results of our simulation studies with artificial data and the empirical application, respectively. Section 7 concludes with some discussion. Details of the MCMC sampling scheme and additional empirical results can be found in the Appendix.

3.2. Time-varying parameter approach for the GMVP

3.2.1. Setup

Let $R_t = (R_{1,t}, ..., R_{n,t})'$ denote a vector of returns on n assets at period t with a covariance matrix Σ. The vector of GMVP weights for the n assets (subject to the constraint that they sum to one) is denoted by $w = (w_1, \ldots, w_n)'$ and is defined as

$$w = \arg\min_{\tilde{w}:\, \iota_n'\tilde{w}=1} \tilde{w}'\Sigma\tilde{w} = \frac{\Sigma^{-1}\iota_n}{\iota_n'\Sigma^{-1}\iota_n}, \qquad (3.1)$$

where ι_n is a $(n \times 1)$ vector of ones. As shown by Kempf and Memmel (2006), the GMVP weights can be represented in terms of the population coefficients in an auxiliary regression in which the return of an arbitrarily selected baseline asset is regressed on a constant and the return differences between the baseline asset and the remaining $n-1$ ones. If asset n is the selected baseline asset, this auxiliary regression is

$$R_{n,t} = \mu + \bar{R}_t'w_{1:n-1} + \epsilon_t, \qquad \mathbb{E}(\epsilon_t|\bar{R}_t) = 0, \qquad (3.2)$$

$$\text{with} \quad w_n = 1 - \iota_{n-1}'w_{1:n-1}, \qquad (3.3)$$

where ϵ_t is the disturbance term and $\bar{R}_t = (R_{n,t} - R_{1,t}, \ldots, R_{n,t} - R_{n-1,t})'$, $w_{1:n-1} = (w_1, \ldots, w_{n-1})'$. The intercept is the expected return of the GMVP, i.e., $\mu = \mathbb{E}(R_t'w)$. The OLS estimates for $w_{1:n-1}$ and w_n in the auxiliary regression (3.2) based on a data series $\{R_t\}_{t=1}^T$ are equal to the 'plug-in' estimates obtained by replacing in Equation (3.1) the population covariance matrix Σ by the sample covariance matrix.

Such an approach to estimating the GMVP weights is known to be problematic for the following reasons: First, when the number of assets n is large relative to the number of observations T, the estimation of Σ by the sample covariance matrix is prone to overfitting due to the excessive number of free parameters. Hence, for an increasing ratio n/T, the accuracy of the sample covariance matrix deteriorates, leading to an increasing bias in the resulting GMVP-weight estimates (see Section 3.1 for a discussion of possible approaches to address this issue).

Second, the OLS approach (and corresponding regularized or shrinked versions thereof) treat the (co)variation among the returns as constant over time. However, there is a plethora of evidence that asset returns tend to exhibit conditional heteroscedasticity which typically implies that also the conditional GMVP weights vary over time. Taking this time variation into account can be expected to be important for GMVP predictions. It should be noted that conditional dynamic approaches in general do not minimize the unconditional portfolio variance, against which investment managers are typically evaluated, but the average conditional variance (law of total variance; see Ferson and Siegel, 2001 for a more detailed discussion). However, as illustrated, e.g., in Reh et al. (2021), since the variation in the conditional covariance structure is typically much more pronounced than the variation in

the conditional expected value in applications to daily stock returns, the optimization of the 'incorrect' target quantity plays a minor role in this case.

3.2.2. Standard GMVP regression with time-varying parameters

A natural way to deal with dynamic changes of GMVP weights while keeping the GMVP analysis robust in high-dimensional applications against overfitting is to treat the GMVP regression model (3.2) as a time-varying parameter (TVP) model and then to combine it with Bayesian shrinkage priors leading to a sparse Bayesian TVP approach, as considered, e.g., by Belmonte et al. (2014) and Bitto and Frühwirth-Schnatter (2019). Applied to the GMVP auxiliary regression model (3.2) we assume for the regressions coefficients random walks of the following form:

$$\begin{bmatrix} \mu_t \\ w_{1:n-1,t} \end{bmatrix} = \begin{bmatrix} \mu_{t-1} \\ w_{1:n-1,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{0,t} \\ \eta_{1:n-1,t} \end{bmatrix}, \qquad \begin{bmatrix} \eta_{0,t} \\ \eta_{1:n-1,t} \end{bmatrix} \sim \mathcal{N}(0, V), \qquad (3.4)$$

where the covariance matrix of the innovations η_t is restricted to be diagonal, i.e., $V = \text{diag}(v_0, v_1, \ldots, v_{n-1})$.

In such a TVP-version of the GMVP regression model, Bayesian shrinkage approaches can be used to shrink the GMVP weights $w_{1:n,t}$ to the weights of some fixed reference portfolio denoted by $\bar{w} = (\bar{w}_1, \ldots, \bar{w}_n)'$. We specifically consider the hierarchical Bayesian LASSO of Belmonte et al. (2014) as well as the related double Gamma prior specification proposed by Bitto and Frühwirth-Schnatter (2019) (which encourages more variation in the shrinkage intensities among the parameters given some global level of shrinkage). If the model is overfitting, these approaches enable an automatic reduction of time-varying weights to constant ones ($v_i \to 0$) and an automatic shrinkage of the weights to those of the reference portfolio ($v_i \to 0$ and $w_{i,t} \to \bar{w}_i$).

While this approach is a way to account for potential dynamic time-variation in the GMVP weights which is robust against large dimensions it suffers from a drawback in that it is not invariant w.r.t. the arbitrary choice of the baseline asset when it is directly applied to the standard GMVP-regression as given by Equation (3.2).

The reason for this lack of invariance is the adding-up constraint in Equation (3.3) leading to an asymmetric treatment of the weights. So this constraint together with the diagonal structure of the covariance matrix V in Equation (3.4), implies for the weight of the baseline asset n a random walk of the form

$$w_{n,t} = w_{n,t-1} + \eta_{n,t}, \qquad\qquad \eta_{n,t} = -\iota'_{n-1}\eta_{1:n-1,t},$$

where

$$\mathbb{E}(\eta_{n,t}^2) = \sum_{i=1}^{n-1} v_i, \qquad \mathbb{E}(\eta_{n,t}\eta_{i,t}) = -v_i, \qquad i = 1, \ldots, n-1.$$

Thus, while all the weights of the first $n-1$ assets are assumed to be mutually uncorrelated, they are all correlated with the weight of the baseline asset n. As a consequence, the diagonal parameterizations of the TVP-GMVP models under different baseline assets are not one-to-one transformations of one another. It follows that the estimation results obtained under different baseline assets are not mutually compatible and lead to different predictions for the GMVP weights.

As is easy to verify, invariance of the TVP-GMVP model w.r.t. the selection of the baseline asset could be achieved by relaxing the restriction on V and allowing it to be an unrestricted non-diagonal covariance matrix. But even in this case, the sparse Bayesian TVP approach applied to the GMVP regression model (3.2) remains problematic. First, if the covariance matrix V is a full matrix, the TVP model is far from being parsimonious from the start, and it is unclear how to design Bayesian shrinkage priors for a full covariance matrix V which could be used to distinguish which of the weights have significant time variation and which do not. Second, even if the TVP-GMVP model with an unrestricted V matrix itself is invariant, the adding-up-constraint generates an invariance issue in the Bayesian analysis of the model. This is because this constraint implies that the prior distribution selected for the covariance matrix V fully determines the prior distribution for the variation parameters of the weight for the baseline asset. So under standard prior assumptions for an unrestricted covariance matrix V (such as a Wishart or Inverse-Wishart) the variation parameters for the baseline asset, $\mathbb{E}(\eta_{n,t}^2)$ and $\mathbb{E}(\eta_{n,t}\eta_{i,t})$, will have prior distributions that differ from the corresponding prior distributions selected for $\mathbb{E}(\eta_{i,t}^2)$ and $\mathbb{E}(\eta_{i,t}\eta_{j,t})$ in V. As a consequence, the resulting posterior distribution for the GMVP weights w_t will depend on the choice of the baseline asset.

3.3. Sparse invariant time-varying parameter model for the GMVP

3.3.1. Augmented GMVP regression

In order to circumvent the invariance problem of the sparse TVP model when it is directly applied to the standard GMVP regression model (3.2) we propose to utilize the GMVP regression which obtains when reformulating the GMVP problem using an artificially augmented asset space that contains a reference portfolio in addition to the n individual assets (Frey and Pohlmeier, 2016). This leads to a reparameterization of the GMVP regression, which enables a symmetric treatment of the GMVP-weights despite their adding-up constraint and thus avoids invariance issues.

Let $R_t^a = (R_{1,t}, \ldots, R_{n,t}, R_{n+1,t})'$ denote the vector of returns in the augmented asset space, where $R_{n+1,t} = \bar{w}' R_t$ is the return of the reference portfolio with fixed weights $\bar{w} = (\bar{w}_1, \ldots, \bar{w}_n)'$ summing to one. The $(n+1)$-dimensional vector of the GMVP weights for this augmented asset space is denoted as $\breve{w} = (\breve{w}_1, \ldots, \breve{w}_n, \breve{w}_{n+1})'$ and is defined by

$$\breve{w} = \arg \min_{\tilde{w}: \, \iota'_{n+1} \tilde{w} = 1} \tilde{w}' \Sigma_a \tilde{w},$$

where Σ_a is the covariance matrix of R_t^a. Since $R_{n+1,t}$ is a linear combination of R_t, this covariance matrix is singular with rank n, so that the GMVP vector $\breve{w}$ is not unique (One of its multiple solutions is $\tilde{w}_{\mathrm{MP}} = \Sigma_a^+ \iota_{n+1} / (\iota'_{n+1} \Sigma_a^+ \iota_{n+1})$, where Σ_a^+ denotes the Moore-Penrose pseudo-inverse of the covariance matrix Σ_a).

Since no further risk reduction can be achieved by expanding the asset space to include the reference portfolio, the resulting variance minimum in the augmented asset space corresponds to that in the regular n-dimensional space. This implies that the GMVP in the augmented space must have the same allocation of the n individual assets as the GMVP in the regular space. From this it follows that the relationship between the original n GMVP weights in the regular space w and the weights for the augmented space $\breve{w}$ is given by the following link function:

$$w = \breve{w}_{n+1} \bar{w} + \breve{w}_{1:n} = (1 - \iota'_n \breve{w}_{1:n}) \bar{w} + \breve{w}_{1:n}, \tag{3.5}$$

where $\breve{w}_{1:n} = (\breve{w}_1, \ldots, \breve{w}_n)'$ (see Frey and Pohlmeier, 2016). Hence, the GMVP weights for the n individual assets in the augmented space $\breve{w}_{1:n}$ represent the (not unique) deviations of their GMVP weights in the regular space w from the scaled weights of the reference portfolio $\bar{w}$. In the link function (3.5), the adding up constraint on $\breve{w}$ and $\bar{w}$ ensures that the GMVP weights w also add up to one. In contrast to the original GMVP weights w, however, the weight deviations for the n assets $\breve{w}_{1:n}$ are not subject to an adding-up constraint. Because of this property of $\breve{w}_{1:n}$ we can treat them symmetrically in a predictive approach in order to ensure a symmetric treatment of the original GMVP weights w via the link function (3.5). This is key for a parametrization of the GMVP regression model for an invariant Bayesian TVP prediction approach.

The optimal weight deviations $\breve{w}_{1:n}$ written in terms of the population coefficients in an auxiliary regression obtain by applying the same regression algebra as Kempf and Memmel (2006). If the reference portfolio is used as the baseline asset, this augmented auxiliary GMVP regression is

$$R_{n+1,t} = \mu + (\bar{R}_t^a)' \breve{w}_{1:n} + \epsilon_t, \tag{3.6}$$

where $\bar{R}_t^a = (R_{n+1,t} - R_{1,t}, \ldots, R_{n+1,t} - R_{n,t})'$. Here the regressors in $\bar{R}_t^a$ are collinear so that the regression coefficients $\breve{w}_{1:n}$ are not identified which reflects the non-uniqueness of

$\breve{w}$. However, this lack of identification in the augmented GMVP regression does not cause any difficulties in a Bayesian analysis as long as the priors are informative with a curvature in all directions where the likelihood as a function in $\breve{w}_{1:n}$ is flat (see, e.g., Bauwens et al., 2000, Section 2.2.4). Rather, this augmented GMVP regression allows us to exploit that the deviations from the GMVP weights $\breve{w}_{1:n}$ for all n assets can be treated symmetrically; and this, in combination with the link function (3.5), enables a Bayesian analysis of the GMVP weights w, which is invariant with regard to the selection of a baseline asset. Moreover, by using Bayesian shrinkage priors that pull the deviations $\breve{w}_{1:n}$ in Equation (3.5) to zero, this augmented approach facilitates an automatic shrinkage of the GMVP weights w to those of the reference portfolio $\bar{w}$, when the model is overfitting. In our primary application below we use the naïve equally weighted portfolio with $\bar{w}_i = 1/n$ as the reference portfolio. This selection follows DeMiguel et al. (2009); Candelon et al. (2012); Frey and Pohlmeier (2016) who consider shrinkage of the GMVP weights towards equality in a static framework.

3.3.2. Augmented GMVP regression with time-varying parameters

The dynamic TVP extension of the augmented auxiliary regression (3.6) we use for a sparse Bayesian GMVP approach is the same TVP model as considered by Bitto and Frühwirth-Schnatter (2019). It is given by

$$y_t = x_t'\beta_t + \epsilon_t, \qquad\qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2), \qquad\qquad (3.7)$$

$$\beta_t = \beta_{t-1} + \omega_t, \qquad\qquad \omega_t \sim \mathcal{N}(0, Q), \qquad\qquad (3.8)$$

where $y_t = R_{n+1,t}$ and $x_t = (1, (\bar{R}_t^a)')'$ with the vector of regression coefficients partitioned as $\beta_t = (\beta_{0,t}, \beta_{1:n,t}')'$ so that $\beta_{0,t} = \mu_t$ and $\beta_{1:n,t} = \breve{w}_{1:n,t}$. The unknown initial value β_0 is assumed to be independent of the innovations $\{\epsilon_t\}$ and $\{\omega_t\}$ and to have a normal prior distribution,

$$\beta_0 \sim \mathcal{N}(\alpha, P_0 Q),$$

where α is an unknown fixed vector and $P_0 = \text{diag}(p_0)$, $p_0 = (p_{0,0}, p_{0,1}, \ldots, p_{0,n})$. For the covariance matrix of the innovations ω_t we use a parsimonious diagonal form with $Q = \text{diag}(q_0, q_1, \ldots, q_n)$, which implies according to Equation (3.8) that the mean GMVP return μ_t and the weight deviations $\breve{w}_{1:n,t}$ are independent random walks.

For given values for β_t, the period-t GMVP weights obtain from the link function (3.5), which can be rewritten as

$$w_t = \bar{w} + R\beta_{1:n,t}, \qquad R = I_n - \bar{w}\iota_n', \qquad\qquad (3.9)$$

where I_n denoted the n-dimensional identity matrix. The first derivatives of w_t with respect

to β_{it}, given by

$$\frac{\mathrm{d}w_t}{\mathrm{d}\beta_{i,t}} = (-\bar{w}_i, \ldots, -\bar{w}_i, (1 - \bar{w}_i), -\bar{w}_i, \ldots, -\bar{w}_i)',$$

$i = 1, \ldots, n$, indicate that an increase of the GMVP weight deviation β_{it} for asset i increases its GMVP weight $w_{i,t}$ by $(1 - \bar{w}_i)$ and decreases the GMVP weights of all the other assets $w_{j,t}$, $j \neq i$ by $\bar{w}_j$, respectively. This represents the mechanics in the link function (3.9) which ensures that the $w_{i,t}$'s obtained from the unrestricted $\beta_{i,t}$'s add up to one. Combining this link function with the independent random walks for $\beta_{1:n,t}$ in Equation (3.8) yields for the n GMVP weights correlated random walks of the form

$$w_t = w_{t-1} + e_t, \qquad\qquad e_t \sim \mathcal{N}(0, RQ_{1:n}R'),$$

where $Q_{1:n} = \mathrm{diag}(q_1, \ldots, q_n)$. Thus, the correlation structure in the innovations e_t is fully determined by the mechanics enforcing the adding-up constraint on w_t.

Since the dependent variable of the GMVP regression in Equation (3.7) is the return on the reference portfolio, it is to be expected that the disturbance term ϵ_t exhibits autoregressive time-varying volatility typical of asset returns. The control for the potential time-variation in the volatility of ϵ_t is of crucial importance, otherwise there is the risk of overestimating the variation in the GMVP weights (Sims, 2001; Stock, 2001).

Although the GMV portfolio weights are chosen in such a way that the idiosyncratic risk of the assets is optimally hedged, it is not possible to diversify away market risk which is faced by the economy as a whole. In the present model, the variance of the error term corresponds to the variance of the portfolio and hence, potential variation in the market risk is captured by heteroscedasticity in the error term. Assuming that the volatility process σ_t^2, $t = 1, \ldots, T$ evolves stochastically as dynamic process with idiosyncratic innovations leads to the class of stochastic volatility models originally proposed by Taylor (1982). This simple and likewise popular approach is to impose for the transition of the log squared volatilities $h_t = \log(\sigma_t^2)$ a stationary AR(1)-process[1] which reads as

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \eta_t^h, \qquad \eta_t^h \overset{iid}{\sim} \mathcal{N}(0, \sigma_h^2), \qquad h_0 \sim \mathcal{N}(\mu_h, \sigma_h^2/(1 - \phi_h^2)). \qquad (3.10)$$

Given that financial data are generally leptokurtic, it might be sensible to impose a Student-t process for the error term. Initial experiments, however, revealed that this leads to negligible differences in the weight forecasts such that we maintain the Gaussian assumption for our models.

---

[1]Initial experiments with alternative SV specifications also yielded satisfactory out-of-sample performance in our empirical applications (see Appendix B.3), but overall no substantial improvements compared to the considered model.

3.4. Bayesian analysis and forecasting

3.4.1. Shrinkage prior specifications

We utilize MCMC methods for Bayesian posterior analysis and use the Gibbs approach to simulate from the joint posterior of the parameters and the latent states. In order to disentangle the initial level and the time-variation in β_t, we rewrite the model in its *non-centered* parametrization, that is, β_t is decomposed into a constant part α and a time-varying part γ_t:

$$\beta_t = \alpha + Q^{0.5}\gamma_t, \tag{3.11}$$

$$\gamma_t = \gamma_{t-1} + \tilde{\omega}_t, \qquad \tilde{\omega}_t \overset{iid}{\sim} \mathcal{N}(0, I_{n+1}), \qquad \gamma_0 \sim \mathcal{N}(0, P_0),$$

where $Q^{0.5} = \text{diag}(\sqrt{q}_0, \ldots, \sqrt{q}_n)$, and the each $\sqrt{q}_i \in \mathbb{R}$ is defined as the positive and negative square root of q_i. This representation facilitates shrinking q_i towards 0, since the usual Inverse-Gamma prior on q_i is bounded away from zero (see, for example, Frühwirth-Schnatter and Wagner, 2010). Here we can equip both, the elements of α and $\sqrt{q} = (\sqrt{q}_0, \ldots, \sqrt{q}_n)$ with hierarchical Gaussian priors. Under the double Gamma prior specification of Bitto and Frühwirth-Schnatter (2019) this reads as

$$\alpha_i \sim \mathcal{N}(0, \tau_i^2), \qquad \tau_i^2 \mid \lambda^2 \sim \mathcal{G}(a^\tau, a^\tau \lambda^2), \qquad \lambda^2 \sim \mathcal{G}(d_{01}, d_{02}), \qquad a^\tau \sim \mathcal{G}(b^\tau, c^\tau),$$

$$\sqrt{q}_i \sim \mathcal{N}(0, \xi_i^2), \qquad \xi_i^2 \mid \kappa^2 \sim \mathcal{G}(a^\xi, a^\xi \kappa^2), \qquad \kappa^2 \sim \mathcal{G}(e_{01}, e_{02}), \qquad a^\xi \sim \mathcal{G}(b^\xi, c^\xi),$$

for all cross-sectional units $i = 1, \ldots, n$. The hierarchical structure allows for an almost complete data driven prior calibration which prevents from cumbrous tuning of the hyperparameters. In contrast to the hierarchical Bayesian LASSO of Belmonte et al. (2014) in which $a^\xi = a^\tau = 1$, also these parameters is assigned a Gamma hyperprior such that for some global level of shrinkage determined by the parameter λ^2 (κ^2), values of a^τ (a^ξ) smaller one increase the variance for τ^2 (ξ^2) leading to more local shrinkage flexibility.

Noteworthy, this increased flexibility comes with the drawback that (few) potentially undesirably large parameter values become more probable. Therefore, we only consider the double Gamma specification for the elements of α and impose for the elements of $\sqrt{q}$ the original hierarchical Bayesian LASSO of Belmonte et al. (2014) fixing $a^\xi = 1$. This ensures a significant amount of shrinkage for the time variation along *all* cross sectional units, which we find to generate less volatile and hence economically more plausible dynamics for the portfolio weights. This hierarchical hybrid shrinkage prior is only imposed for the parameters corresponding to the augmented portfolio weights, whereas we select a fairly uninformative normal prior for the parameters α_0 and $\sqrt{q}_0$, associated to μ_t, the conditional expectation of the GMVP.

The joint posterior for our proposed TVP-GMVP-*shr*-SV model obtains as

$$\pi\left(\gamma_{0:T}, h_{0:T}, \alpha, \sqrt{q}, \tau^2, \xi^2, \lambda^2, \kappa^2, a^\tau, \mu_h, \phi_h, \sigma_h^2, p_0 \mid y_{1:T}\right),$$

The list of parameters is given by $\theta = (\alpha, \sqrt{q}, \tau^2, \xi^2, \lambda^2, \kappa^2, a^\tau, \mu_h, \phi_h, \sigma_h^2, p_0)$. Exploiting the conjugacy to the Gaussian likelihood, conditional posteriors are mostly available in closed form.

3.4.2. MCMC algorithm

To carry out Bayesian inference for the model defined in Equations (3.7), (3.10) and (3.11), we mostly adapt the MCMC sampling scheme of Bitto and Frühwirth-Schnatter (2019) for a sparse TVP model with hierarchical shrinkage priors, including an interweaving step to improve the efficiency of the sampler.[2] For the stochastic volatility, we utilize the standard approach of Kim et al. (1998) to draw the volatility states h and associated parameters. As suggested by initial investigations, the hyperparameters for ϕ_h are adjusted to obtain a highly informative prior enforcing high persistence of the volatility states for the sake of out-of-sample forecasting performance. For our forecasting experiments, additional to our *shrinkage (shr)* models we will present all results also for a *no shrinkage (no shr)* specification, in which the same single-layer Gaussian priors as for α_0 and $\sqrt{q}_0$ are imposed for all elements of α and $\sqrt{q}$. Moreover, we implement additional to the stochastic volatility specification (SV) also a homoscedastic variant of our model (no SV). In the no SV-model, the states h and the associated parameters of the volatility process are replaced by a time-invariant volatility parameter σ^2, which is equipped with a hierarchical Gamma prior. For all details on prior selection, see the Appendix B.1. For a detailed description of the MCMC sampling scheme see Appendix B.2.1.

3.4.3. Forecasting

After dropping the draws from the first cycles as burn-in we use the draws from the next S cycles for the purpose of approximating the joint augmented posterior $\pi(\theta, \gamma_{0:T}, h_{0:T} \mid y_{1:T})$. Bayesian point estimates (posterior means) of the model parameters are then obtained as sample averages over the corresponding Gibbs draws.[3] Forecasts for some period-t portfolio weight w_t are obtained as deterministic function of β_t (see Equation (3.9)), which is itself recovered from the parameters and latent states of the non-centered notation as $\beta_t = \alpha + Q^{0.5}\gamma_t$. For predictive performance comparisons with alternative non-Bayesian forecasting approaches we rely on one-step-ahead point forecasts of the weights β_{T+1} for given values of

---

[2] The model is implemented in Matlab 2018a. To sample from the GIG distribution, we use a Matlab implementation of the method proposed in by Hörmann and Leydold (2014), which is available on `https://de.mathworks.com/matlabcentral/fileexchange/78805-gigrnd`.

[3] In our simulations to artificial data (Section 3.5) as well as in the empirical application (Section 3.6) we run the MCMC algorithm for parameter estimation for 15,000 iterations, where the first 5000 are discarded.

the parameters θ, i.e., we use the conditional expected value as one-step-ahead point forecast $\hat{\beta}_{T+1}$:

$$\hat{\beta}_{T+1} := \mathbb{E}(\beta_{T+1} \mid y_{1:T}, \theta) = \mathbb{E}(\alpha + Q^{0.5}\gamma_{T+1} \mid y_{1:T}, \theta).$$

This value is approximated by

$$\widehat{\mathbb{E}(\beta_{T+1} \mid y_{1:T}}, \theta) = \alpha + Q^{0.5}\,\widehat{\mathbb{E}(\gamma_{T+1} \mid y_{1:T}}, \theta)$$
$$= \alpha + Q^{0.5}\frac{1}{S}\left(\sum_{i=1}^{S}\tilde{\gamma}_{T+1}^{(i)}\right),$$

where in the SV specifications $\tilde{\gamma}_{T+1}^{(i)}$ is a draw from the one-step-ahead-predictive distribution $f(\gamma_{T+1} \mid y_{1:T}, h_{1:T}^{(i-1)}, \theta)$ from the Kalman filter in the reduced Gibbs sampler for h and γ, and θ is fixed at its posterior mean.[4] In the homoscedastic specifications $\mathbb{E}(\gamma_{T+1} \mid y_{1:T}, \theta)$ obtains from the one-step-ahead-predictive distribution $f(\gamma_{T+1} \mid y_{1:T}, \theta)$ from the Kalman filter, where θ including the volatility σ^2 is fixed at its posterior mean.

3.5. Monte Carlo simulations

In this section, we examine and compare the performance of our sparse TVP-GMVP regression with that of various benchmark strategies in an in-sample and an out-of-sample simulation exercise. A particular focus of this experiment is to analyze the ability of the proposed GMVP models to track the 'true' conditional GMVP weights. We follow the simulation study of Engle et al. (2019) and sample from a DCC with realistic parameter settings for daily returns.

3.5.1. Baseline scenario

The DGP reads as

$$R_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_t),$$

with $\mathcal{F}_{t-1}$ denoting the information set up to period $t-1$. The conditional covariance matrix Σ_t evolves according to the DCC-GARCH model of Engle (2002). It is decomposed into

$$\Sigma_t = D_t^{0.5}C_tD_t^{0.5},$$

where C_t is the conditional correlation matrix and $D_t = \text{diag}(h_{1t}, \ldots, h_{nt})$ is a diagonal matrix with the conditional return variances, each of which follows a univariate GARCH(1,1)

---

[4]In our out-of-sample experiments with artificial data (Section 3.5) as well as in the empirical application (Section 3.6) we run the reduced Gibbs sampler for 4000 iterations, where the first 1500 are discarded.

process. The correlation matrix C_t is given by

$$C_t = (Q_t^*)^{-0.5} Q_t (Q_t^*)^{-0.5},$$

with $Q_t = (1 - \tilde{\alpha} - \tilde{\beta})S + \tilde{\alpha} e_{t-1} e_{t-1}' + \tilde{\beta} Q_{t-1}$, where $e_t = D_t^{-0.5} R_t$ is the vector of standardized returns. The diagonal matrix Q_t^* is composed of the diagonal elements of Q_t, and S is the unconditional covariance matrix of the standardized returns. In order to implement correlation targeting, S is estimated by the sample covariance of the standardized returns e_t that are based on univariate GARCH models. We set the correlation parameters to $\tilde{\alpha} = 0.05$, $\tilde{\beta} = 0.93$, the parameters for the univariate volatility dynamics to $a_i = 0.05, b_i = 0.90 \; \forall i$ and the unconditional population variance is set to the unconditional empirical covariance matrix of the n most liquid stocks of our data set using 10 years of daily data from 2005 to 2014 (see Section 3.6.1, for details on the data set).

We construct portfolio weights by plugging the covariance Σ_t (computed in period $t-1$) into the GMVP formula given in Equation (3.1) which leads to the following *true* time $t-1$ conditional GMVP weights:

$$w_t^* = \frac{\Sigma_t^{-1} \iota_n}{\iota_n' \Sigma_t^{-1} \iota_n},$$

such that the GMVP is given as $R_{pt} := w_t^{*\prime} R_t$ with conditional variance $\sigma_{pt}^2 := w_t^{*\prime} \Sigma_t w_t^*$.

3.5.2. Competing models

As alternatives to our proposed GMVP models we use several approaches that are motivated by their popularity in the literature. In order to address the estimation noise associated with the estimation of high-dimensional covariance matrices, whenever applicable, shrinkage techniques are applied also to these benchmark models, which are briefly described in the following. For the approaches that predict (conditional) covariances Σ (Σ_t), we construct the period-t prediction for the GMVP weights by plugging the covariance prediction $\hat{\Sigma}$ ($\hat{\Sigma}_t$) into the GMVP formula given in Equation (3.1).

(i) The standard DCC as outlined above as well as its nonlinear shrinkage version (DCC-nl) which is developed and recommended by Engle et al. (2019) who consider it the 'new DCC standard in large dimensions'. The DCC-nl modifies the original DCC model by using the nonlinear shrinkage estimator of Ledoit and Wolf (2012, 2015) – instead of the sample covariance matrix – for correlation targeting. Both DCC models are implemented using the assumption of normally distributed errors and are estimated by a composite likelihood approach as recommended by Engle et al. (2019).

(ii) A regularized exponential Recursive Least Squares with forgetting factor scheme (RLS-REF) to directly infer the GMVP weights (Reh et al., 2021), based on a linear regression of a benchmark return on the return differences to this benchmark, in which the

regression coefficients represent the GMVP weights (Kempf and Memmel, 2006). The forgetting factor $\lambda \in (0, 1]$ operates as an exponential weight decreasing for more remote observations. It is treated as an unknown parameter to be estimated consistently by minimizing the expectation of the quadratic loss function

$$L(\beta_t, R_t) = (Y_t - X_t'\beta_t)^2,$$

whereby the sequence of parameters $\{\beta_t\}$ obtain from the recursion

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \Omega_t^{-1} X_t (Y_t - X_t'\hat{\beta}_{t-1}), \tag{3.12}$$

$$\Omega_t = X_t X_t' + \lambda \Omega_{t-1} + (1 - \lambda)\Omega_0, \tag{3.13}$$

so that $\hat{\beta}_{1:n-1t}$ and $\hat{\beta}_{0t}$ are the estimates for the period-t GMVP weights and mean return with some initial conditions (β_1, Ω_0). β_1 is set equal to the vector $(0, \iota_{n-1}/n)'$ implying a conditional portfolio mean of zero and portfolio weights all equal to $1/n$. The information matrix Ω_0 is derived from an equicorrelation-equivariance matrix for the return vector, which is scaled by an additional model parameter that determines the degree of shrinkage towards the equally weighted portfolio defined by β_1 and Ω_0.

(iii) A Wishart multivariate stochastic volatility model (WSV) for the conditional precision matrix denoted H_t^{-1}. Following Uhlig (1994, 1997), the evolution of H_t^{-1} is governed by a singular multivariate Beta distribution shock as follows:

$$H_t^{-1} = \frac{d+1}{d}\mathcal{U}\left(H_{t-1}^{-1}\right)'\Theta_t\mathcal{U}\left(H_{t-1}^{-1}\right),$$

$$H_1^{-1} \sim \mathcal{W}_n\left(d, [dS_0]^{-1}\right),$$

where $\mathcal{U}\left(H_t^{-1}\right)$ is the upper triangular matrix obtained from the Cholesky decomposition of H_t^{-1}. The shocks Θ_t are iid draws from an n-dimensional singular multivariate Beta distribution $\mathcal{B}_n\left(\frac{d}{2}, \frac{1}{2}\right)$ as defined by Uhlig (1994), with $d > n - 1$ degrees of freedom, $\mathcal{W}_n$ denotes the n-dimensional Wishart distribution and the initial covariance matrix is given as $S_0^{-1} = \mathbb{E}\left[H_1^{-1}\right]$.

Uhlig (1997) shows that the nonlinear filtering of the latent precision matrices can be computed analytically. The predictive density of the precision matrix is given by

$$p\left(H_{t+1}^{-1} \mid R_t\right) \sim \mathcal{W}_n\left(d, [dS_{t+1}]^{-1}\right)$$

where S_{t+1} evolves according to

$$S_{t+1} = \frac{d}{d+1}S_t + \frac{1}{d+1}R_tR_t'. \tag{3.14}$$

The dynamics of the precision matrix are governed by a unique parameter, d, that can

be estimated by Maximum Likelihood. Substituting backwards in Equation (3.14) it is possible to obtain the following expression for S_t:

$$S_t = \lambda^t S_0 + (1 - \lambda) \sum_{i=1}^{t-1} \lambda^{i-1} R_{t-i} R'_{t-i}, \tag{3.15}$$

where $\lambda = \frac{d}{d+1} < 1$. Hence, given that $d > n + 1$, large n implies $\lambda \approx 1$ such that the estimated conditional covariance matrices are shrunk towards S_0. In this case, it is crucial to chose an adequate initial covariance matrix. Following Moura et al. (2020), we set the initial condition S_0 in (3.15) to an equicorrelation covariance matrix in which, as outlined in Engle and Kelly (2012), the correlation between any two returns is equal to the average sample correlation between all returns in the portfolio. Further, we denote as the Shrunk WSV (SWSV) a setting in which S_0 is a diagonal matrix whose diagonal elements are given by the in-sample variance of each return, which implies that the one-step-ahead forecasts of the correlation are shrunk towards zero.

(iv) Several static plug-in approaches based on estimates of the unconditional covariance matrix Σ. The OLS estimator constructs the period-t prediction of the GMVP weights by replacing the return covariance matrix with the sample covariance matrix of the returns observed up to period $t-1$. This approach is equivalent to running Kempf and Memmel's (2006) static auxiliary regression using OLS. The linear shrinkage (SHR-l) estimator modifies the OLS estimator by estimating Σ via the linear shrinkage estimator of Ledoit and Wolf (2004). This estimate shrinks the sample covariance matrix towards the identity matrix. It minimizes the expected Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix. Finally, the nonlinear shrinkage (SHR-nl) estimator is based on the nonlinear shrinkage procedure of Ledoit and Wolf (2012, 2015).

(v) The naïve estimator sets the prediction of the GMVP weights equal to the weights of the equally weighted portfolio.

3.5.3. Evaluation

As our approach is tailored to make predictions for GMV portfolio weights, we directly evaluate the forecasts $\hat{w}_t$ as well as the estimated portfolio variance, instead of applying covariance based loss function as suggested, e.g., in Engle et al. (2019):

- GMVP weight forecasts using L_1 penalty (as suggested in Callot et al., 2019, Theorem 3):

$$L_{1,w_t} = ||\hat{w}_t - w_t^*||_1$$

- GMVP weight forecasts using L_2 penalty:

$$L_{2,w_t} = ||\hat{w}_t - w_t^*||_2$$

- Portfolio variance:

$$\hat{\sigma}_{pt}^2 = \hat{w}_t'\Sigma_t\hat{w}_t.$$

As the simulation study is designed such that we sample from a (conditionally) mean zero process, it holds that $\mathbb{V}(R_{pt}) = \mathbb{E}[\mathbb{V}(R_{pt} \mid \mathcal{F}_{t-1})]$, such that for an unconditional evaluation we can compare average volatilities given by

$$\sigma_V^2 = \frac{1}{T}\sum_{t=1}^T \sigma_{pt}^2.$$

Furthermore, we define the average losses for the weights as $L_{i,w} = \frac{1}{T}\sum_{t=1}^T L_{i,w_t}$, $i = 1,2$.[5]

In order to compare the weight forecasts of our TVP-GMVP models to those obtained by observation driven approaches like the DCC, we need a filtration for the weights in our in-sample experiment, i.e., $\hat{w}_t \mid \mathcal{F}_{t-1}, \hat{\theta}$. In case σ_t^2 follows itself some stochastic process $\{h_t\}$, we need to integrate out the volatility states numerically. For that we make use of a marginalized bootstrap particle filter (with 25,000 particles) for $\{h_t\}$ using Rao-Blackwellization based on the Kalman filter for $\{\gamma_t\}$ (see Schön et al., 2005). For the specifications without SV, $\hat{w}_t \mid \mathcal{F}_{t-1}, \hat{\theta}$ is readily available by the Kalman filter. Further details on the filtration for the portfolio weights are deferred to Appendix B.2.2.

3.5.4. Results

To evaluate the in-sample performance of our proposed models in comparison to the correctly specified DCC and the remaining benchmark approaches, we simulate $n_S = 500$ data sets and compute for all models the series of filtered GMVP weights. We consider different number of assets $n = \{100, 200\}$ and time series of different length $T = \{250, 1250\}$. Table 3.1 reports the losses averaged over time and all data sets. For assessing the statistical significance of differences in losses, we apply the model confidence set (MCS) approach of Hansen et al. (2011) taking the time averaged values σ_V^2, $L_{1,w}$ and $L_{2,w}$ as loss series. Based on the maximal t-statistic for the pairwise loss differentials of all models under consideration, the MCS is constructed to contain the best-performing models at a given confidence level, for which we consider 75% and 90%.[6] The results are extremely persistent from one simulation

---

[5]For the out-of-sample experiment T is replaced by the number of out-of-sample forecasts denoted T_s.

[6]In the bootstrap implementation of the MCS, we use a block bootstrap with block length $\lfloor T_{eval}^{1/3} \rfloor$, where T_{eval} is equal to the size of the evaluation sample, and a bootstrap sample size of 10000. We use the implementation of the MCS procedure in the Oxford MFE toolbox (`https://www.kevinsheppard.com/code/matlab/mfe-toolbox`).

to the next which is indicated by low variation in the losses across the data sets. Thence, the MCS contain at most two model specifications although the relative differences in the average losses are fairly small in several settings.

Among our proposed specifications, the models with hierarchical shrinkage priors outperform the *no shr* specifications in all settings and for all losses. Likewise, this holds among the standard compared to the respective shrinkage specifications of the benchmark models. Overall, this improvement is, as expected, more pronounced for higher concentration ratios n/T. For none of the settings, static approaches are included in the MCS for any of the losses.

Notably, although the data generating process is a DCC, our TVP-GMVP-*shr* models significantly outperform the DCC(-nl) for $n = 100$ and $n = 200$ under the shorter estimation window. In the most challenging situation with $n/T = 0.8$, the lowest variance is generated for the no SV-*shr* model whereas the average losses with respect to the weight predictions are lowest for SV-*shr* indicating more regularization of the weight dynamics when accounting for potential systematic heteroscedasticity. For $n = 200$, $T = 1250$, although the DCC-nl generates the lowest average portfolio variance, the best weight predictions are again obtained by the TVP-GMVP-SV-*shr*-model. Presumably, the flexible shrinkage priors regularize the portfolio weights such that the model is less sensitive to outliers. This feature is even more pronounced for the SV specifications in which is controlled for heteroscedasticity in the market, which may otherwise be incorrectly reflected in the weight dynamics. However, this robustness is of minor relevance for in-sample evaluations and can be expected to improve the forecasting performance more substantially in the out-of-sample exercise as well as in our empirical application. Solely for the lowest concentration ratio, the DCC-nl outperforms all other specifications significantly.

In Table 3.2, we report the average losses of an out-of-sample simulation exercise. Here we simulated $n_S = 100$ data sets of length $T + T_s$, $T_s = 100$. All models are re-estimated for all time periods in a rolling window scheme, such that in total, each $T_s \times n_S = 10,000$ one-step ahead predictions are evaluated. Again, we rely on the MCS approach of Hansen et al. (2011), taking the time averaged losses for all simulations as loss series. The results are qualitatively fairly similar to the in-sample exercise and again, we observe substantial improvements under the shrinkage models. Notably, out-of-sample the increased potential regularization of the portfolio weights of the SV models in comparison to the homoscedastic specifications indeed appears to become more beneficial for GMVP prediction. Except for $\hat{\sigma}_V^2$ for $n = 100$, $T = 1250$, the TVP-GMVP-SV-shr specification is included in all 75% MCS and leads to the lowest average loss in 8 out of 12 comparisons. Among the benchmarks, additional to the DCC-nl, also the predictions generated from the static SHR-nl approach as well as the (S)WSV model lead to low $L_{1,w}$ and $L_{2,w}$ losses, particularly for the short estimation window. For the latter, the parameter estimates (not reported here) indicate a high degree of smoothing of the conditional covariances. While this may be too restrictive

Table 3.1.: In-sample simulation results

Model	$\hat{\sigma}_V^2$	$\hat{L}_{1,w}$	$\hat{L}_{2,w}$	$\hat{\sigma}_V^2$	$\hat{L}_{1,w}$	$\hat{L}_{2,w}$
	$n = 100, T = 250$			$n = 100, T = 1250$		
true DGP	0.114	0	0	0.115	0	0
TVP-GMVP-						
SV-*shr*	0.172	2.986	0.390	0.168	3.086	0.396
SV-*no shr*	0.227	5.080	0.621	0.170	3.394	0.436
no SV-*shr*	**0.167**	**2.939**	**0.387**	0.169	3.035	0.387
no SV-*no shr*	0.230	5.083	0.617	0.168	3.299	0.424
DCC	0.237	5.398	0.797	0.153	3.188	0.398
DCC-nl	0.179	3.500	0.490	**0.148**	**2.876**	**0.361**
RLS-REF	0.189	3.157	0.405	0.170	3.113	0.398
WSV	0.275	3.296	0.438	0.201	3.343	0.432
SWSV	0.265	3.179	0.414	0.198	3.288	0.428
naïve $(\frac{1}{n})$	0.998	3.996	0.529	1.000	4.001	0.528
OLS	0.248	5.379	0.671	0.251	3.631	0.443
SHR-l	0.225	4.603	0.583	0.250	3.606	0.440
SHR-nl	0.197	3.367	0.436	0.246	3.350	0.408
	$n = 200, T = 250$			$n = 200, T = 1250$		
true DGP	0.068	0	0	0.068	0	0
TVP-GMVP-						
SV-*shr*	0.126	**4.261**	**0.408**	0.110	**4.032**	**0.389**
SV-*no shr*	0.212	6.858	0.624	0.116	4.698	0.440
no SV-*shr*	**0.1207**	4.314	0.413	0.127	4.326	0.405
no SV-*no shr*	0.194	6.696	0.610	0.127	4.642	0.434
DCC	0.335	12.021	1.120	0.110	4.943	0.464
DCC-nl	0.127	4.575	0.434	**0.100**	4.153	0.390
RLS-REF	0.151	4.599	0.439	0.119	4.195	0.391
WSV	0.293	4.827	0.479	0.153	4.271	0.408
SWSV	0.288	4.720	0.466	0.148	4.086	0.396
naïve $(\frac{1}{n})$	1.061	5.311	0.519	1.059	5.306	0.519
OLS	0.365	13.101	1.219	0.157	5.063	0.469
SHR-l	0.167	6.626	0.602	0.156	4.980	0.461
SHR-nl	0.133	4.552	0.431	0.150	4.298	0.401

Note: Mean average loss L_w and mean average volatility σ_V^2 for $n_S = 500$ simulations ($\hat{\sigma}_V^2 = \frac{1}{n_S} \sum_{k=1}^{n_S} \sigma_V^{2(k)}$, $\hat{L}_{1,w} = \frac{1}{n_S} \sum_{k=1}^{n_S} L_{1,w}^{(k)}$, $\hat{L}_{2,w} = \frac{1}{n_S} \sum_{k=1}^{n_S} L_{2,w}^{(k)}$). Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS, using $\sigma_V^{2(k)}$, $L_{1,w}^{(k)}$, $L_{2,w}^{(k)}$, $s = 1, \ldots 500$ as loss series.

to capture relevant dynamics in the correlation structure and hence the GMVP weights in-sample, the disadvantage of reduced flexibility appears to be outweighed by the noise reduction in out-of-sample forecasting.

Table 3.2.: Out-of-sample simulation results

Model	$\hat{\sigma}_V^2$	$\hat{L}_{1,w}$	$\hat{L}_{2,w}$	$\hat{\sigma}_V^2$	$\hat{L}_{1,w}$	$\hat{L}_{2,w}$
	$n = 100, T = 250$			$n = 100, T = 1250$		
true DGP	0.1149	0	0	0.115	0	0
TVP-GMVP-						
SV-*shr*	**0.160**	**2.958**	**0.390**	0.165	2.996	0.394
SV-*no shr*	0.209	4.833	0.630	0.164	3.132	0.412
no SV-*shr*	0.176	3.610	0.475	0.240	3.513	0.459
no SV-*no shr*	0.207	4.823	0.630	0.254	3.611	0.471
DCC	0.238	5.392	0.718	0.161	3.283	0.432
DCC-nl	0.183	3.568	0.478	**0.155**	**2.975**	**0.391**
RLS-REF	0.169	3.220	0.409	0.170	3.199	0.407
WSV	0.204	3.053	0.399	0.193	3.783	0.493
SWSV	0.206	2.980	0.393	0.192	3.814	0.498
naïve ($\frac{1}{n}$)	1.014	3.998	0.532	1.003	4.011	0.538
OLS	0.244	5.395	0.704	0.275	3.747	0.489
SHR-l	0.221	4.611	0.594	0.275	3.721	0.485
SHR-nl	0.198	3.504	0.447	0.270	3.479	0.452
	$n = 200, T = 250$			$n = 200, T = 1250$		
true DGP	0.068	0	0	0.068	0	0
TVP-GMVP-						
SV-*shr*	**0.107**	**4.035**	**0.384**	0.104	**3.816**	**0.359**
SV-*no shr*	0.674	15.281	1.289	0.108	4.211	0.395
no SV-*shr*	0.115	4.410	0.417	0.141	4.504	0.419
no SV-*no shr*	0.654	14.201	1.282	0.137	4.647	0.433
DCC	0.316	11.771	1.108	0.113	4.941	0.468
DCC-nl	0.126	4.556	0.439	**0.103**	4.159	0.394
RLS-REF	0.119	4.533	0.421	0.109	3.975	0.362
WSV	0.198	4.632	0.457	0.121	4.180	0.385
SWSV	0.211	4.512	0.441	0.118	4.137	0.384
naïve ($\frac{1}{n}$)	1.082	5.319	0.523	1.053	5.303	0.521
OLS	0.342	13.030	1.215	0.165	5.057	0.470
SHR-l	0.156	6.467	0.587	0.164	4.976	0.461
SHR-nl	0.126	4.595	0.425	0.159	4.374	0.403

Note: Mean average loss L_w and mean average volatility σ_V^2 for $n_S = 100$ simulations each with $T_S = 100$ re-estimations in a rolling window scheme producing $T_S = 100$ one-step-ahead forecasts ($\hat{\sigma}_V^2 = \frac{1}{n_S} \sum_{k=1}^{n_S} \sigma_V^{2(k)}$, $\hat{L}_{1,w} = \frac{1}{n_S} \sum_{k=1}^{n_S} L_{1,w}^{(k)}$, $\hat{L}_{2,w} = \frac{1}{n_S} \sum_{k=1}^{n_S} L_{2,w}^{(k)}$). Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS, using $\sigma_V^{2(k)}$, $L_{1,w}^{(k)}$, $L_{2,w}^{(k)}$, $s = 1, \ldots 100$ as loss series.

3.6. Empirical application

The goal of this section is to examine the out-of-sample properties of our newly suggested approach for modeling portfolio weights using actual data. Additionally to out-of-sample

variances, we present out-of-sample Sharpe ratios for these investment strategies under realistic trading costs and compare it to full Markowitz portfolios based on the covariance estimates of the benchmark models.

3.6.1. Data and set-up

The data base is the same as used in Moura et al. (2020) and consists of prices of *all* NYSE, AMEX and NASDAQ stocks observed daily from 01/02/2002 to 11/21/2019. The models are recursively estimated using a rolling window scheme based on investment universes with $n \in (100, 200, 400)$ assets each for a one year ($T = 250$) and a five year ($T = 1250$) estimation period, which allows us to evaluate the performance of our suggested models from a concentration ratio $n/T = 0.08$ up to a challenging ratio $n/T = 1.6$.[7]

The investment universe is obtained as follows: We find the set of stocks that have a complete return history over the initial estimation period as well as the complete out-of-sample window. We then look for possible pairs of highly correlated stocks, that is, pairs of stocks with returns with a sample correlation exceeding 0.95 over the first estimation period. With such pairs, if they should exist, we remove the stock with the lower volume on the last year of the first estimation period denoted h. Of the remaining set of stocks, we then pick the largest n stocks (as measured by their market capitalization on the investment date h) as our investment universe. The parameters are re-estimated every month adopting the common convention that 21 consecutive days constitute one month. Hence for each specification we perform a total of 155 rolling window estimations to obtain an out-of-sample period of 13 years starting from 12/18/2006 to 11/21/2019 with a total number of $S = 3255$ out-of-sample periods. All portfolios are updated monthly to avoid an unreasonable amount of turnover and thus transaction costs.

3.6.2. GMVP variance

In this section, we present the empirical out-of-sample portfolio variance $\hat{\sigma}^2$ of our dynamic GMVP models as well as the dynamic and static benchmark approaches. For some sequence of portfolio return R_{ps}, $s = 1, \ldots, S$, with empirical mean $\bar{R}_p = \frac{1}{S} \sum_{s=1}^{S} R_{ps}$, it is computed as

$$\hat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^{S} (R_{ps} - \bar{R}_p)^2.$$

For the benchmark models for which we obtain point estimates for the (conditional) covariance, the GMVP weights are obtained as plug-in estimates according to Equation (3.1). For assessing the statistical significance of differences in the empirical out-of-sample variance, we

---

[7]A typical setting for empirical analysis of daily financial asset returns is $T = 1250$ (see e.g., Engle et al., 2019; Callot et al., 2019; Moura et al., 2020).

Table 3.3.: Out-of-sample GMVP variance with $T = 250$

	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-			
SV-*shr*	0.523	**0.485**	0.405
SV-*no shr*	0.773	1.556	3.581
no SV-*shr*	0.591	0.552	**0.387**
no SV-*no shr*	0.802	1.152	4.583
DCC	0.756	1.156	—
DCC-nl	0.623	0.559	0.483
RLS-REF	0.548	0.492	0.428
WSV	0.535	0.516	0.515
SWSV	**0.521**	0.488	0.498
naïve ($\frac{1}{n}$)	1.411	1.450	1.669
OLS	0.687	1.436	—
SHR-l	0.645	0.867	0.624
SHR-nl	0.537	0.487	0.415

Note: Out-of-sample model comparison. Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS.

again apply the MCS approach of Hansen et al. (2011) (see Section 3.5.4), taking now the demeaned squared returns as loss series.

Table 3.3 reports the results for an estimation window of one year ($T = 250$ observations) where concentration ratios vary from 0.4 to 1.6, with more cross sectional units than time periods for $n = 400$. We observe that the shrinkage specifications of our models overall clearly outperform the specifications with standard non-hierarchical priors, although also the *no shr* models have zero centered priors and therefore deliver moderate shrinkage in direction of the (time-stable) equally weighted portfolio. Despite the short estimation period, the specification which takes into account potential heteroscedasticity in the market portfolio leads to significantly lower portfolio variance when compared to the homoscedastic specification for $n = 100$ and $n = 200$. For the setting with $n > T$, the estimation uncertainty due to the increased number of parameters appears to outweigh the benefits of accounting for stochastic volatility. The best performance among the benchmark models in this setting is observed for SHR-nl and the sparsely parameterized dynamic specifications, namely, RLS-REF, WSV and particularly SWSV, which are the only competitors to be included in the 90% model confidence set for $n = 100$ and $n = 200$. However, in the most challenging setting with $n > T$, only TVP-GMVP-no SV-*shr* is included in the 90% and 75% MCS and all benchmarks perform substantially worse than our proposed shrinkage approaches. The out-of-sample variance of SWSW is even larger than that of the portfolios generated from the nonlinear shrinkage versions of DCC and the unconditional covariance estimate. This leads to the conclusion that our dynamic GMVP shrinkage model, and in particular, the stochastic volatility specification thereof, is the most robust among all considered models.

Table 3.4.: Out-of-sample GMVP variance with $T = 1250$

	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-			
SV-*shr*	**0.520**	0.463	**0.394**
SV-*no shr*	0.537	0.497	0.464
no SV-*shr*	0.554	0.488	0.412
no SV-*no shr*	0.584	0.517	0.469
DCC	0.611	0.540	0.507
DCC-nl	0.599	0.521	0.464
RLS-REF	0.534	0.485	0.409
WSV	0.541	0.465	0.408
SWSV	0.541	**0.457**	0.397
naïve $\left(\frac{1}{n}\right)$	1.350	1.417	1.587
OLS	0.556	0.500	0.479
SHR-l	0.555	0.499	0.474
SHR-nl	0.549	0.484	0.419

Note: Out-of-sample model comparison. Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS.

This finding is confirmed by the results of the five year estimation period, which we report in Table 3.4. The TVP-GMVP-SV-*shr* is the only model to be included in the 90% MCS for all cross sectional dimensions. Taking into account potential heteroscedasticity is here found to be beneficial overall. As expected, the advantage of the shrinkage specifications is less pronounced in this setting, also among the benchmark models. Furthermore, the dynamic models outperform the static approaches more clearly.

3.6.3. Mean-variance efficiency

Now we turn to the classical analysis of mean-variance efficiency of the portfolios generated by our proposed models compared to those of the benchmarks. For evaluating the performance, we use the empirical Sharpe ratio defined as

$$\hat{\theta} = \frac{\bar{R}_p}{\sqrt{\hat{\sigma}^2}},$$

with a superior forecasting performance indicated by larger values of the Sharpe ratio. In order to take into account the costs of rebalancing, we additionally report the Sharpe ratio net of transaction costs. Following e.g., Kirby and Ostdiek (2012) and Moura et al. (2020), we compute the turnover adjusted portfolio return as

$$R_{ps}^{\text{adj.}} = (1 - c \, \text{turnover}_s)(1 + R_{ps}) - 1, \quad \text{turnover}_s = \sum_{i=1}^{n}(|w_{js} - w_{js}^*|),$$

where w_s^* denotes the allocation vector at period $s-1$ after taking into account the changes in asset prices between periods $s-1$ and s and c is the fee that must be paid for each transaction, which is measured in terms of basis points (bp.). Referring to the discussion in French (2008) who observed a 92% decrease in the costs of trading from the 1980's to approximately 11 basis points in 2006, we set $c=10$, taking into account that our out-of-sample period comprises the time span 2006 – 2019.

Given that we forecast the GMVP weights directly, the allocations remain the same as in the previous section for our proposed models and the RLS-REF, as well as trivially, also for the naïve diversification. For all other benchmarks, we consider a mean-variance portfolio based on an investor who aims at minimizing the portfolio risk subject to a target portfolio return. Hence, the weights are the solution on the following optimization problem for a given covariance estimate $\hat{\Sigma}_s$

$$\operatorname*{argmin}_{w_s} w_s'\hat{\Sigma}_s w_s, \quad \text{s.t. } w_s'\iota_n = 1, w_s'm = \mu_{\text{targ}}. \tag{3.16}$$

Various approaches exist to construct the signal m and to choose the target return μ_{targ}. We follow the same approach as Engle et al. (2019) and Moura et al. (2020) and construct the signal using the *momentum* factor of Jegadeesh and Titman (1993). For each of the n stocks the individual momentum is computed as the geometric average of the previous 252 returns, but excluding the 21 most recent returns. Collecting all the momentums in a vector yields the signal m. The target return is computed as the arithmetic average of the momentums of those stocks that belong to the top-quintile stocks ranked according to momentum. The analytical solution to (3.16) is given by

$$w_s = \hat{\Sigma}_s^{-1}\frac{m(Cb-D)+\iota_n(E-Db)}{EC-D^2},$$

with $C = \iota_n'\hat{\Sigma}_s^{-1}\iota_n$, $D = m'\hat{\Sigma}_s^{-1}\iota_n$ and $E = m'\hat{\Sigma}_s^{-1}m$. One could argue that the insights from a comparison of different investment strategies (such as GMVP vs. mean-variance with momentum optimization) are limited. However, what we want to illustrate here in particular is that GMV portfolios perform well even when the optimization problem and associated evaluation criterion deviates from variance minimization.

The results for $T=250$ displayed in Table 3.5 are very similar to those obtained for the GMVP variance estimation. Among our models, the shrinkage prior leads to substantially larger Sharpe ratios and the stochastic volatility specifications are advantageous for all but the setting with concentration ratio larger than one. Furthermore, we observe that along with the static SHR-nl estimate, the RLS-REF and the (S)WSV models are again the best competitors. Presumably, their particularly good performance for the short estimation window can be traced back to their sparsity (only one parameter is estimated in the (S)WSV and two parameters are estimated in the RLS-REF model). Moreover, the estimates for d (not reported here) in the S(WSV) specifications imply a smoothing parameter close to one

Table 3.5.: Out-of-sample Sharpe ratio ($\times 10$) with $T = 250$

	$c = 0$ bp.			$c = 10$ bp.		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-						
SV-*shr*	0.605	**0.611**	0.686	0.530	0.494	0.570
SV-*no shr*	0.484	0.279	0.308	0.324	-0.072	-0.184
no SV-*shr*	0.647	0.586	**0.710**	0.0535	0.432	**0.621**
no SV-*no shr*	0.481	0.337	0.224	0.312	0.109	0.183
DCC	0.450	0.249	—	0.197	-0.220	—
DCC-nl	0.520	0.585	0.603	0.345	0.385	0.390
RLS-REF	0.647	0.608	0.649	0.587	0.561	0.539
WSV	**0.682**	0.605	0.487	**0.622**	**0.559**	0.438
SWSV	0.651	0.600	0.539	0.588	0.542	0.486
naïve ($\frac{1}{n}$)	0.490	0.462	0.450	0.482	0.451	0.442
OLS	0.521	0.337	—	0.367	-0.061	—
SHR-l	0.556	0.409	0.482	0.415	0.122	0.186
SHR-nl	0.655	0.576	0.526	0.563	0.474	0.413

Note: *Mean-variance with a momentum signal* weights for DCC(-nl) and static specifications, GMVP weights for TVP regression and RLS-REF specification. Largest value in bold letters. Results are displayed for $c = 0$ and $c = 10$ basis points trading costs.

in all estimation windows ($\hat{d} > 1000$) imposing much weight on the starting value S_0 as well as a smooth evolution of the portfolio weights associated with moderate turnover.

Finally, in Table 3.6 we report the results for the five year estimation period. Our shrinkage models lead to the highest Sharpe ratios for all n without transaction costs, and in 2/3 settings with transaction costs. The increased reduction of the Sharpe ratios after accounting for turnover under *no shr* priors demonstrates the importance of dynamic sparsity induced by the LASSO type priors for the elements of α and $\sqrt{q}$. Regularizing portfolio weights leads to more stable allocations associated with lower transaction costs. Notably, the homoscedastic shrinkage model is better than all competitor models except for the other direct dynamic GMVP forecasting approach RLS-REF, which performs slightly better in two settings. This suggests that, although the GMVP is conceptually not optimal in terms of Sharpe ratio maximization, the noise reduction due to the direct targeting of the portfolio weights enables a good overall performance.

Summing up, we advocate the use of the TVP-GMVP model with double Gamma and Bayesian LASSO type priors for initial level and dynamics of the portfolio weights. It is the only model that leads to good out-of-sample results for all considered (n, T)-specifications providing allocations that lead to low variance and high Sharpe ratios. Its flexibility allows to be competitive under a five year estimation window in which models that account for the dynamic in the correlation structure of the data outperform static approaches, as well as under the one year estimation window in which sparsity is most crucial to prevent from the curse of dimensionality. Accounting for conditional heteroscedasticity in the market

Table 3.6.: Out-of-sample Sharpe ratio ($\times 10$) with $T = 1250$

	$c = 0$ bp.			$c = 10$ bp.		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-						
SV-*shr*	0.589	0.614	**0.716**	0.516	0.519	**0.580**
SV-*no shr*	0.610	0.600	0.648	0.519	0.480	0.473
no SV-*shr*	**0.618**	**0.666**	0.705	0.533	**0.566**	0.558
no SV-*no shr*	0.616	0.657	0.685	0.515	0.0535	0.496
DCC	0.506	0.561	0.586	0.285	0.0276	0.203
DCC-nl	0.515	0.574	0.601	0.309	0.0325	0.303
RLS-REF	0.610	0.628	0.706	**0.566**	0.559	0.519
WSV	0.605	0.589	0.597	0.507	0.496	0.512
SWSV	0.611	0.604	0.631	0.511	0.511	0.542
naïve ($\frac{1}{n}$)	0.457	0.446	0.454	0.438	0.428	0.419
OLS	0.529	0.525	0.497	0.461	0.434	0.343
SHR-l	0.532	0.530	0.508	0.466	0.441	0.358
SHR-nl	0.538	0.565	0.575	0.475	0.491	0.486

Note: *Mean-variance with a momentum signal* weights for DCC(-nl) and static specifications, GMVP weights for TVP regression and RLS-REF specification. Largest value in bold letters. Results are displayed for $c = 0$ and $c = 10$ basis points trading costs.

is particularly relevant when evaluating solely the GMVP variance, whereas for a mean-variance optimizing investor, also the sparser specification with homoscedastic errors is well suitable.

3.6.4. Extensions

Dynamic benchmark allocations

So far, we have used the equally weighted portfolio composed from weights $\bar{w}_i = 1/n$ as the reference portfolio. This leads to the desired regularization by the Bayesian shrinkage priors and, as shown, good out-of-sample performance. However, the equally weighted portfolio performs as allocation rule significantly worse than the other models in all our experiments, raising the question of whether our TVP-GMVP model could also perform better under a different reference portfolio. One advantage of a more informative benchmark allocation could be that a revision of the prior would only be necessary if the data show patterns in the correlation structure that cannot be captured by the underlying benchmark model. As this regards both the overall level as well as the dynamics in the implied GMVP weights, in this section we analyze the performance of our models based on a dynamic benchmark portfolio, that is, a benchmark return that results from an allocation that varies over time. We model the deviations from this dynamic benchmark, where the shrinkage prior ensures that there is only a deviation from this allocation if there is substantial evidence in the data. The mostly data-driven calibration of the hyperparameters via hierarchical priors remains as

in our original model specification (see Section 3.4.1). To ensure that our prior continues to provide regularization that prevents overfitting and counteracts large in-sample estimation errors, only benchmark allocations that exhibit moderate volatility over time and do not have extremely large exposures are sensible to use. Moreover, a dynamic benchmark as allocation rule should itself deliver satisfactory results in GMVP comparisons.
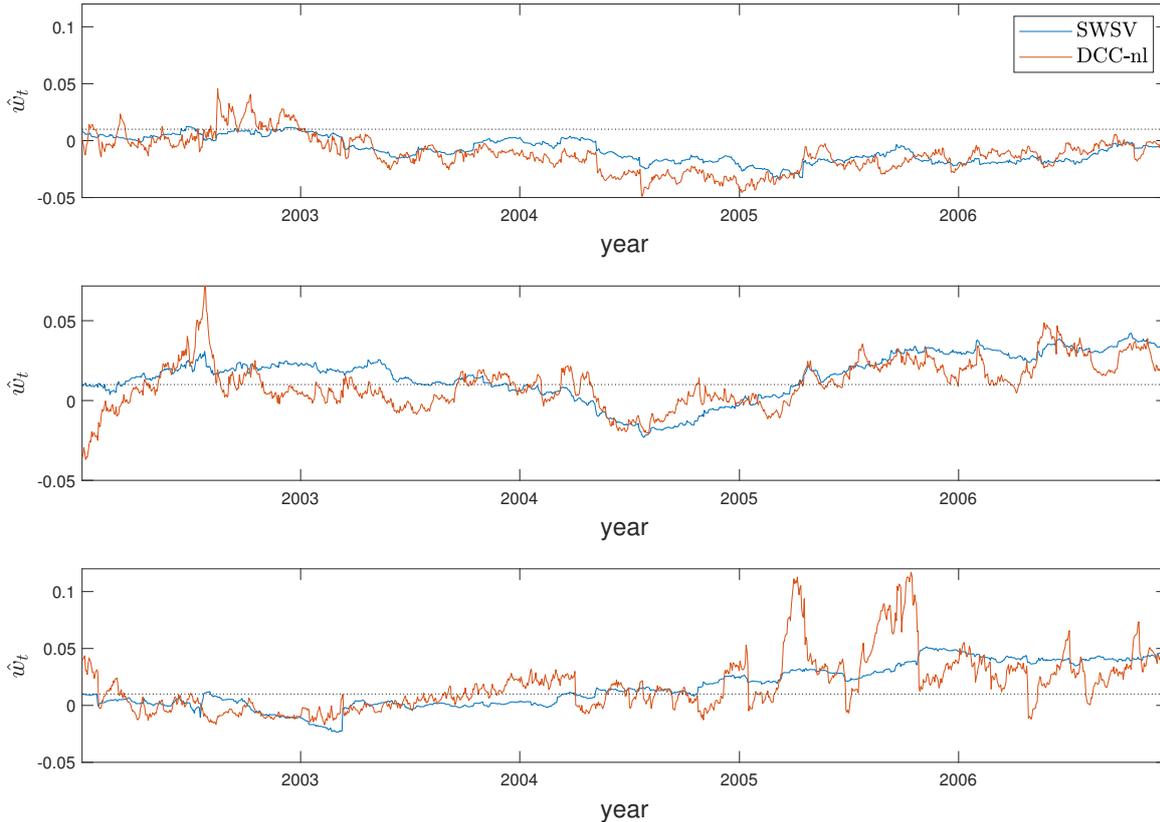
Figure 3.1.: In-sample GMVP weight estimates $\hat{w}_t$ for first estimation period (time span 01/02/2002 – 12/15/2006) under $(n, T) = (100, 1250)$ compared to $1/n$ (black dotted line).

The GMVP allocations which are derived from the Wishart multivariate stochastic volatility models satisfy these requirements. Particularly, in our empirical analysis, we found that the shrinkage version, SWSV, performed very well in terms of minimum variance allocations. Figure 3.1 shows exemplarily for three randomly selected stocks for our first estimation period (time span 01/02/2002 – 12/15/2006) the evolution of the in-sample fitted SWSV-implied GMVP weights compared to those implied by the DCC-nl model for the specification $(n, T) = (100, 1250)$. The evolution of the SWSV weights is much smoother and tends to exhibits less gross exposure. Yet, all weight series have some time variation and clearly deviate from the naïve allocation (here $1/n = 0.01$, indicated by black dotted line). This indicates that using the SWSV-implied GMVP weights instead of the naïve benchmark will result in the shrinkage prior being less restrictive, which implies that less revision of the prior assumptions is needed to broadly follow the trend of the GMV allocation. Although the

SWSV model is clearly also unable to identify the *true* path of the desired portfolio weights, the simulation and empirical results suggest that the model is closer to the GMVP than the equally weighted allocation (likewise, the predicted conditional covariances are closer to the true conditional covariance than those implying the equally weighted portfolio, like, e.g., multiples of the identity matrix or equivariance-equicorrelation matrices). In situations where the covariance estimates imply strong re-balancing, under this dynamic benchmark allocation this would be automatically built in our TVP-GMVP model without the need for the corresponding parameters to have high variation, i.e., the elements of $\sqrt{q}$ to become very large.

Table 3.7 shows the resulting empirical portfolio variances of the TVP-GMVP-*shr* specifications for the out-of-sample forecasting experiment under a benchmark return $R_{n+1,t} = \bar{w}_t' R_t$, where we set $\bar{w}_t = \hat{w}_t^{SWSV}$, $t = 1, \ldots, T$, the in-sample predictions of the GMVP weights resulting from the SWSV model and $\bar{w}_{T+1} = \hat{w}_{T+1}^{SWSV}$, the corresponding one-step-ahead out-of-sample prediction. From this it follows that the relationship between the original n GMVP weights in the regular space w_t and the weights for the augmented space $\breve{w}_t$ at some period t is given by the following link function (see Equation (3.5)):

$$w_t = \breve{w}_{n+1,t} \hat{w}_t^{SWSV} + \breve{w}_{1:n,t} = (1 - \iota_n' \breve{w}_{1:n,t}) \hat{w}_t^{SWSV} + \breve{w}_{1:n,t}.$$

Since the specification with this dynamic benchmark is in some sense based on insights from the out-of-sample experiment, we consider the results only complementary to our main empirical analysis, not including them in the model confidence set comparisons presented in Section 3.6.2. In fact, the results reveal that the use of an informative benchmark is very promising: The alternative specifications in 9/12 settings (along all n, T, and each with and without stochastic volatility) lead to lower portfolio variances than the corresponding model with the original benchmark $1/n$. In particular, the results in the model without stochastic volatility are substantially improved for all n and T except for $(n, T) = (400, 250)$. This can be explained by the fact that in this scenario the SWSV model is itself not competitive, exhibiting almost 30% higher estimated GMVP variance than the original TVP-GMVP-no SV-*shr* model.

Extreme concentration ratios

Our empirical application has shown that the utility of shrinkage priors is high especially in situations in which the number of cross sectional units is large compared to the sample size with the prior contribution to the posterior comparably large. As a second complementary analysis, we now present results for scenarios with even more challenging concentration ratios in order to robustify our findings on the benefits of the shrinkage priors.

In doing so, we examine the performances of our shrinkage models compared to those of the parsimonious or shrinkage variants of the benchmarks, both when T decreases and when n increases. Particularly, we consider the scenarios $(n, T) = (800, 250)$ and $(n, T) = (400, 125)$

Table 3.7.: Out-of-sample GMVP variance with dynamic SWSV benchmark return

	$T = 250$		
	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-			
SV-SWSV-*shr*	0.536	**0.483**	0.407
no SV-SWSV-*shr*	**0.563**	**0.545**	0.389
	$T = 1250$		
	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-			
SV-SWSV-*shr*	**0.516**	**0.460**	**0.392**
no SV-SWSV-*shr*	**0.535**	**0.474**	**0.404**

Note: Out-of-sample model comparison. Bold value indicates that the variance is lower than respective counterpart with the naïve benchmark.

with concentration ratio $n/T = 3.2$, as well as $(n, T) = (1600, 250)$ and $(n, T) = (400, 63)$ with concentration ratio $n/T \approx 6.4$. The data base and selection of the asset universe is the same as described in Section 3.6.1. Since situations with portfolios composed of more than 1000 stocks of individual companies or a database of only approximately three months of daily data is rather artificial, we consider also this analysis only supplementary to our main empirical application. Of course, it is worth mentioning that such settings may be realistic for other types of assets.

Table 3.8 presents the resulting empirical portfolio variances for our TVP-GMVP-*shr* specifications compared to the applying benchmarks. We observe that in these scenarios, only our shrinkage TVP-GMVP models are included in 75% and 90% MCS. With increasing concentration ratio, both induced by large n or small T, the relative performance of our models, particularly the stochastic volatility specifications, compared to all benchmarks, improves. Among these, the direct GMVP weight modeling approach RLS-REF is the only dynamic specification that tends to slightly outperform the extremely robust static SHR-nl estimates in most scenarios. Interestingly, although the S(WSV) models are not competitive here, the dynamic SWSV-benchmark (which is again not part of the MCS analysis) leads to substantially lower empirical variances than the respective version with static benchmark weights. This let assume that in situations with challenging concentration ratio informative prior weights that capture at least the broad trend in the portfolio weights are particularly useful. They help to control in-sample biases and avoid erroneous overestimation especially for the parameters steering the time-variation in the latent state dynamics. Indeed, throughout all (n, T) specifications, the median and average posterior mean of $\sqrt{q}$ along the cross sectional units (not reported here) are larger under the static weight benchmark compared to the dynamic SWSV benchmark, both for the models with and without heteroscedasticity in the market explicitly taken into account.

The results in this section support our previous finding that the prior-induced shrinkage

Table 3.8.: Out-of-sample GMVP variance for extreme concentration ratios

	$n = 800$ $T = 250$	$n = 1600$ $T = 250$	$n = 400$ $T = 125$	$n = 400$ $T = 63$
TVP-GMVP-				
SV-*shr*	**0.314**	**0.202**	**0.411**	**0.456**
SV-SWSV-*shr*	0.300	0.202	0.405	0.432
no SV-*shr*	0.318	0.212	0.419	0.471
no SV-SWSV-*shr*	0.299	0.199	0.403	0.437
DCC-nl	0.416	0.227	0.517	0.517
RLS-REF	0.358	0.214	0.439	0.483
WSV	0.534	0.443	0.622	0.762
SWSV	0.500	0.357	0.575	0.660
naïve ($\frac{1}{n}$)	1.961	1.855	1.669	1.669
SHR-l	0.388	0.223	0.504	0.505
SHR-nl	0.352	0.219	0.445	0.488

Note: Out-of-sample model comparison. Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS (TVP-GMVP models with SWSW benchmark are not included in the comparisons).

works particularly well in situations with critical concentration ratios. From a technical perspective, it should be added that for our approach the sampling of the latent states γ is the bottleneck in terms of computational effort.[8] High concentration rates, i.e., large n/T, in turn, are computationally not problematic in themselves. For example, for $(n, T) = (800, 250)$, the computational effort is comparable to that for the $(n, T) = (200, 1250)$ scenario which further strengthens the argument that our proposed approach is particularly advantageous in scenarios with large n/T but moderate nT.

3.7. Summary and discussion

We use a time-varying parameter regression to estimate the weights of the Global Minimum Variance Portfolio (GMVP). Different from plug-in approaches which use estimates of the covariance matrix to compute GMVP weights, the proposed method scales linearly and allows to shrink portfolio weights directly, avoiding extreme allocations and encouraging stable portfolios.

In a simulation study based on a DCC data generating process, we show that the TVP-GMVP models lead to satisfactory results and deliver portfolios with smaller variance and more accurate weight forecasts than the DCC as long as $n/T > 0.4$. These conclusions hold for simulations made both in-sample and out-of-sample. The importance of shrinkage

---

[8] The required memory scales quadratically with the term nT when using AWOL or a precision sampler, as here a matrix of dimension $(n + 1)T \times (n + 1)T$ is created. Using FFBS based on the Kalman filter requires less memory, but becomes very slow if the term nT gets large.

is highlighted in the out-of-sample results of the challenging scenario $(n, T) = (200, 250)$, in which the TVP-GMVP model with shrinkage delivers portfolios with one fifth of the risk of the portfolios based on the TVP-GMVP without shrinkage.

In our main empirical exercise based on data sets with up to 400 assets and several competing models tailored for high-dimensional applications, the TVP-GMVP models confirm the results of the simulation study and are shown to deliver portfolios with the lowest variance in 4 out of 6 scenarios, and are always included in the 90% model confidence set. Similar results are obtained when we evaluate Sharpe ratios of the resulting portfolios compared to mean-variance optimal allocations for the considered benchmark models.

As indicated in the simulation study, also the empirical results confirm that the shrinkage priors are particularly useful under challenging concentration rates. To demonstrate the robustness of this finding, in an additional experiment we fit the model with estimation window truncated down to $T = 63$ as well as with the cross-sectional dimension increased up to $n = 1600$. Indeed, in both situations (T decrease or n increase) the performance improvements become more evident compared to the benchmark models considered.

For future analysis, specifications with alternative, potentially dynamic, benchmark allocations could be investigated in more detail which may lead to further improvements of the predictive performance of the proposed TVP-GMVP approach, as indicated by the promising results in Section 3.6.4.

It has to be acknowledged that although the shrinkage prior can be used to prevent extreme allocations with high short-selling shares and large turnover, our model in its current form is limited to the unrestricted GMVP, unlike conventional covariance-based plug-in approaches. However, in applications to daily returns, there is much evidence in the literature (see, e.g., DeMiguel et al., 2009) that the GMVP can be superior to other mean-variance optimal allocations in Sharpe ratio comparisons, which is also suggested by the results of our empirical application in Section 3.6.3. Yet, the limitation to the unrestricted GMVP also implies that explicit constraints, e.g., on large-exposure, cannot be taken into account. In that regards, Zhao et al. (2021) have recently shown that gross-exposure constraints are mathematically equivalent to the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2015), which, however, has the advantage that instead of one exogenous penalty parameter, it has n degrees of freedom that are inherently determined in the optimization. The authors conclude that nonlinear shrinkage is superior to setting gross-exposure constraints in out-of-sample experiments, which they further demonstrate in simulations and empirical applications. Similarly, in our approach, it can be argued that the data-driven shrinkage procedure induced solely via prior choice is superior to setting specific hard constraints, at least in situations in which the restriction on allocations serves only as a means to the end of risk reduction. In practice, however, there are also situations in which the asset manager is subject to certain restrictions in her investment decisions. An approach to explicitly incorporate gross-exposure constraints or other restrictions on the weights could be the following:

First, run the Gibbs sampler to simulate from the joint posterior of the parameters and latent states determining the portfolio weights, potentially using special reference portfolios (see Section 3.6.4) or tighter calibrations of the prior distributions than in our current specifications. Second, for out-of-sample forecasting of portfolio weights satisfying a certain restriction, simulate from the truncated predictive density satisfying the restriction at hand.

A second drawback of our model is the computational effort. For example, the MCMC sampler to infer the parameters on a computer with Xeon 3.70 GHz processor with 10 cores (Intel Xeon W-2255) requires about 168 minutes in the setting $(n, T) = (100, 1250)$ and about 45 hours in the setting $(n, T) = (400, 1250)$ for $15,000$ iterations. For the stochastic volatility specifications there comes added the computational time of the reduced Gibbs sampler to integrate out the latent volatility states for generating out-of-sample weight predictions. However, having the parameter re-estimation and portfolio reallocation occur only monthly, as is typically done, this issue is not too problematic. For practicability, one might consider having the re-estimation of parameters occur even less frequently.

Chapter 4.

Inferring Dynamic Financial Networks via a Time-Varying Graphical LASSO Approach with Applications to Portfolio Selection

4.1. Introduction

Modeling of financial portfolio weights is based on the interdependencies among the return data. Except for simple allocations such as the equally weighted portfolio, weights are usually selected as an optimal basket out of a set of assets based on how they are interrelated. In particular, mean-variance optimal strategies as derived from Markowitz' (1952) portfolio theory like the Sharpe ratio maximizing (maxSR) allocation and the global minimum variance portfolio (GMVP) for the special case of an infinitely risk averse investor, are obtained as scaled linear functions of the *inverse* of the joint covariance matrix of the asset returns, the so-called precision matrix. In this work, I propose to model financial precision matrices directly, without the detour on the covariance. I exploit the scaled linear linkage between the elements of the precision matrix and portfolio weights, which facilitates interpretation and allows specific requirements such as regularizing the weights to reduce estimation noise to be addressed during the precision estimation process.

In the recent years, high dimensional portfolio selection based on financial asset returns with hundred or more cross-sectional units has become increasingly popular due to higher data availability and increased computational power. The major challenge of modeling the covariation is that the number of elements in the covariance or its inverse increases quadratically in the number of assets which can lead to critical concentration ratios, i.e., situations in which the number of cross-sectional units is high compared to the number of estimation periods. This makes all approaches prone to the curse of dimensionality, including static estimates of the unconditional covariance as well as dynamic approaches which take into account potential heteroscedasticity in the data. Many strategies exist to address the high-dimension problem, though each has limitations: First, with regard to the dynamic models, sparse parameterizations like the scalar specification for the correlation dynamics in the dynamic conditional correlation (DCC) model of Engle (2002) are essential to make estimation

feasible in high dimensional applications. However, they may be too restrictive and fail to describe the dynamics of the conditional covariances appropriately. Second, factor models help to capture only relevant structures in the covariation (Ledoit and Wolf, 2003; Fan et al., 2013; De Nard et al., 2021; Lee and Seregina, 2021), but it is not always clear what the factors are, or factor data may not be available. Third, shrinkage approaches of the unconditional covariance or correlation matrix are generally designed for improving the bias-variance trade-off of the estimation (Ledoit and Wolf, 2020): Under linear shrinkage techniques (Ledoit and Wolf, 2003, 2004), the estimates obtain as a linear combination of the sample covariance and a target matrix like a multiple of the identity matrix, a factor model or an equicorrelation covariance; under nonlinear shrinkage (Ledoit and Wolf, 2012, 2015), the regularization reduces dispersion of the eigenvalues and assigns an individual shrinkage intensity to each eigenvalue of the covariance estimate. Despite the advance of shrinkage techniques in the recent years, these approaches remain to some extent suboptimal for portfolio allocations because the determination of the optimal shrinkage intensities relies on minimizing certain loss functions that are not directly related to optimal portfolio allocation. Alternatively to (regularized) covariance or precision estimation, it is also possible to model portfolio weights directly from the return observations, e.g., based on linear regression representations (Britten-Jones, 1999; Kempf and Memmel, 2006). A drawback of these approaches is that they are limited to a particular allocation rule, say GMVP predictions. However, recent results have proven that they perform very well in applications tailored to their specific problem (Frey and Pohlmeier, 2016; Ao et al., 2019; Reh et al., 2021) suggesting that it may be advantageous to consider the quantity to be optimized already when estimating or calibrating a model. For similar reasons, Bodnar et al. (2021) recently proposed penalizing deviations from previous weight forecasts in rolling window type out-of-sample experiments. This sparsity in dynamics is aimed at the economically important turnover costs, on one hand, and is useful from a statistical point of view to reduce estimation noise, on the other.

Precision matrix modeling allows to combine the advantages of flexibility with respect to particular allocations, such as strategies with limited gross exposure, that covariance models offer, with the advantage of the targeted perspective of direct weight modeling. While existing work on precision modeling has particularly emphasized the advantages of avoiding the inversion of potentially ill-conditioned covariance estimates in applications with high concentration ratios (see, e.g., Callot et al., 2019; Caner et al., 2020), I focus on the scaled linear association of the elements of the precision matrix with portfolio weights. Conceptually, the precision matrix maps the conditional dependencies of the assets, suggesting an interpretation as an undirected graph. To reduce estimation noise, an obvious assumption is to impose sparsity constraints on the graph via LASSO (least-absolute shrinkage and selection operator)-like penalty terms, which is potentially more plausible from an economic perspective for precision matrices than for covariances of financial asset returns: While a zero entry in the covariance matrix, corresponding to a zero correlation between two assets, is very

unlikely in highly interconnected financial systems typically analyzed in this context, a zero entry in the precision in a Gaussian model requires conditional independence, which may be plausible especially in the presence of highly correlated industries. However, under a factor structure, for example, the conditional independence assumption may also be too harsh, so this paper proposes to include penalty functions for the precision that allow regulation of certain portfolio allocations without requiring a sparse structure of the graph.

Based on the time-varying graphical LASSO (TVGL) of Hallac et al. (2017), I develop a sparse dynamic Gaussian graphical model for the precision matrix in which I assume a piecewise constant joint distribution for the returns. Unlike previously developed graphical models for precision matrices in financial applications (see, e.g., Janková and van de Geer, 2018; Callot et al., 2019; Lee and Seregina, 2021), the TVGL is not limited to a static estimate of the unconditional inverse covariance. Moreover, it can be tailored to specific problems, as I show using a newly introduced penalty function that regularizes the weights of the global minimum variance portfolio. This penalty is designed to reduce the gross exposure of GMVP weights without enforcing sparsity in covariance or precision, allowing, for example, equicorrelation-like dependencies between assets that are exposed in classical LASSO. Because of the combination of penalties on the precision matrix and its dynamics, the augmented TVGL model (ATVGL) is a flexible mixture between a time-stable and a dynamic model. It allows for temporal variations in the precision and the inclusion of conditional information, but differs conceptually from other dynamic approaches that model a sequence of conditional covariance matrices. In addition, the sliced interpretation of time allows the optimization problem to be blocked, resulting in significant computational savings in the optimization procedure.

Via comparisons of existing portfolio selection strategies under realistic scenarios, the paper additionally contributes to the financial portfolio modeling literature by showing that relative predictive performances exhibit substantial temporal instabilities over long time horizons, suggesting that it may be optimal to dynamically switch, for example, between more robust and more flexible specifications to achieve uniformly good performance. In particular, I show that based on a comparison of existing successful allocation strategies, in a rolling-window forecasting setup, a general superiority of dynamic over static models cannot be empirically demonstrated, and that the length of the estimation window, which is usually not a focus in econometric papers, should be carefully chosen when estimating an unconditional covariance matrix. These insights motivate the development of my proposed forecasting approach: to answer the empirical question of what level of sparsity, temporal stability, and conditioning information is optimal for forecasting, I develop a dynamic recalibration scheme for the penalty parameters that selects the best model in terms of its predictive performance as a function of the current data set and the quantity to be optimized. The ATVGL, which is equipped with a problem-oriented dynamic calibration, selects the conditionally optimal parameterization and thereby implicitly the degree of time stability

and the weighting of the historical data information. Moreover, it allows a fast adaptation to current economic conditions. In an empirical application to daily U.S. traded stock returns, I show that the proposed approach is able to outperform the static and dynamic benchmark models with respect to minimum variance and mean-variance optimal predictions both on average over the 40-year period considered and in each 5-year subperiod.

Recent work has considered applications with up to $n = 2000$ cross-sectional units (Wang et al., 2020). Applications with a very limited number of observation periods T leading to situations with $n > T$ have also been analyzed (Callot et al., 2019; Caner et al., 2020). Although theoretically appealing, from a practitioner's point of view, it is questionable whether allocating among assets from thousands of individual stocks is desirable. To create realistic scenarios, in my analysis I consider portfolio selection based on a large but not vast cross-sectional dimension of the asset universe, i.e., $n \in (100, 200)$. Moreover, since I work with daily stock returns for which historical time series are usually readily available, I refrain from assuming particularly short estimation windows. It is worth mentioning that my proposed approach is neither theoretically nor computationally constrained in terms of the cross-sectional dimension or the concentration ratio.

The remainder of this paper is organized as follows: Based on theoretical and practical insights, Section 2 analyzes the performance of static and dynamic approaches in minimizing out-of-sample unconditional portfolio variance. Section 3 introduces the augmented time-varying graphical LASSO problem and its optimization. Section 4 illustrates my dynamic recalibration scheme and presents empirical results, and Section 5 concludes. Detailed derivations and additional empirical results are deferred to the Appendix.

4.2. Optimal portfolio allocations

4.2.1. Theoretical considerations

The idea of this paper is to develop, on the basis of a graphical model for precision matrices, a model and an associated prediction procedure capable of forecasting portfolio weights that compose optimal portfolios with respect to some evaluation criterion of interest. To this end, in this section I briefly summarize what the notion of optimality means in this context. Let $R_t = (R_{1t}, \dots, R_{nt})'$ denote a vector of n risky financial asset returns at period t. For given joint unconditional covariance $\Sigma = \mathbb{V}[R_t]$ and unconditional mean return $\mu = \mathbb{E}[R_t]$, the vector of mean-variance efficient portfolio weights w^* is given as

$$w^* = \operatorname*{argmin}_{w} w'\Sigma w, \quad \text{s.t.} \tag{4.1}$$

$$\iota'w = 1, \tag{4.2}$$

$$\mu'w = \mu^*, \tag{4.3}$$

where ι is an $n \times 1$ vector of ones and μ^* is a target return chosen according to the investor's risk preferences. This includes infinite risk aversion which leads to the global minimum variance portfolio for which condition (4.3) can be dropped such that the estimation of the expected portfolio return is avoided. The well-known analytic 'plug-in' solution for the GMVP weights as a function of the (inverse) joint covariance matrix is

$$w^{GMV} = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}. \tag{4.4}$$

In practice, Σ and, if applicable, μ typically need to be estimated and μ^* carefully calibrated.

To account for conditional heteroscedasticity which is a common feature in financial return data, dynamic conditional approaches are used: They typically predict the conditional covariance $\Sigma_t = \mathbb{V}[R_t \mid \mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1}$ is the amount of information known at time $t-1$ and the problem defined in Equations (4.1) – (4.3) is solved at each t for the $\mathcal{F}_{t-1}$-conditionally optimal portfolio weights w_t^*. However, since investors are usually evaluated on the basis of the unconditional quantities, a trade-off is created when comparing static unconditional and dynamic conditional approaches by the fact that in the latter conditional information is explicitly taken into account, which in principle can be assumed to be beneficial for the forecasts. On the other hand, a sequence of conditionally optimal forecasts is generated, the average of which in general does not correspond to the actual target, *unconditionally* optimal quantities.[1] For example, for the unconditional variance, this can easily be illustrated by the law of total variance:

$$\Sigma = \mathbb{E}[\Sigma_t] + \mathbb{V}[\mu_t]. \tag{4.5}$$

The potential superiority of dynamic conditional approaches over static ones for the unconditional covariance depends on whether the variation in Σ_t is substantially more pronounced than the variation in the conditional mean return μ_t. As shown, e.g., by Reh et al. (2021), this is typically the case for daily asset returns that are subject to a high degree of conditional heteroscedasticity, near-zero expectations, and an overall poor signal-to-noise ratio. What these theoretical considerations do not take into account, however, is that in practice estimation uncertainty plays a major role. It is particularly pronounced in dynamic models, which tend to be more flexible. This is only one reason why an evaluation of the two model classes based only on the underlying data generating process of the returns is not sufficient. Moreover, in out-of-sample forecasting experiments, re-estimation is typically performed at regular intervals, for example monthly, with the data base evolving over a rolling window. Especially when considering longer time periods, the resulting sequences of estimates for the unconditional covariance show considerable differences, which means that the time windows typically used (five years is fairly standard) are not sufficient to make reliable inference for

---

[1]Note that, as outlined in Ferson and Siegel (2001) models for such mean-variance optimal weights that optimize the unconditional quantities while at the same time explicitly taking into account conditioning information cannot be operationalized in general, and hence are not readily available in practice.

the true underlying stationary covariance matrix – if it exists at all. However, this need not be to the detriment of portfolio forecasting performance. In fact, another trade-off arises especially for the estimation of unconditional covariances: with shorter windows, the estimates are more sensitive and can be interpreted as including time-conditional information, while long estimation windows use a larger data base, which would be advantageous for the inference of a stationary covariance and is associated with lower estimation noise. As extreme example, consider a zero-mean process combined with a discrete Markov conditional volatility process with two possible states (say, a crisis state and a non-crisis state), specified as follows:

$$\mathbb{V}[R_t \mid Z_t] = Z_t \quad \text{with}$$
$$P(Z_t = \Sigma^A \mid Z_{t-1} = \Sigma^A) = 0.999, \quad P(Z_t = \Sigma^B \mid Z_{t-1} = \Sigma^A) = 0.001,$$
$$P(Z_t = \Sigma^A \mid Z_{t-1} = \Sigma^B) = 0.001, \quad P(Z_t = \Sigma^A \mid Z_{t-1} = \Sigma^B) = 0.999.$$

For some sample $R_t, t = 1, \ldots T^S$ sufficiently short, due to the high persistence in the process, it is likely that Z_t takes only one of the values Σ^A or Σ^B (e.g., for $T^S = 250$, the probability is approx. 78%) such that any conventional estimation of the unconditional covariance like the sample covariance would be close to this value, although the unconditional variance of R_t is given as $\mathbb{V}[R_t] = 0.5(\Sigma^A + \Sigma^B)$. For optimal portfolio forecasts, however, it might even be advantageous in this example to choose rather short estimation windows in order to be able to capture the 'local stability' at hand.

In practice, the variation in forecasts and their dependence on the chosen time window clearly depends on many factors. In order to empirically test the relevance of the discussed trade-offs for portfolio allocations based on daily stock returns and what implications for a forecasting model can be drawn from them, I already present an excerpt from my empirical application in the following section. Particularly, I investigate the long-term relative forecasting ability of each one of the most widely used static and dynamic approaches to covariance modeling in recent years.

4.2.2. Empirical out-of-sample comparison of static and dynamic approaches

As a starting point, I consider the work of Moura et al. (2020), which surveys an extensive list of portfolio selection strategies over a time horizon of more than 40 years. For all cross-sectional dimensions, the best average results were obtained with dynamic covariance models. In particular, the DCC model with nonlinear shrinkage of the correlation matrix (DCC-nl; Engle et al., 2019) performed best for the cross-sectional dimensions $n = 100$ and $n = 500$. Among the static approaches for covariance estimation, the nonlinear shrinkage estimation (SHR-nl) of Ledoit and Wolf (2012, 2015) resulted in the best overall out-of-sample predictions for all evaluation criteria. Therefore, I choose these models as exemplary representatives of the dynamic and static approaches to portfolio selection for illustrating

various aspects of portfolio performance over time that are not revealed by the usual inspection of average quantities, which may be strongly influenced by outliers, for example. I generate one-step-ahead covariance forecasts based on the $n = 100$ largest assets from my data set, with assets selected from the same large pool of assets as in the paper by Moura et al. (2020). Based on a rolling window of $T = 1250$ trading days, the model re-estimation and reallocation takes place every 21 trading days, which corresponds to one month (see Section 4.4.1 for details on the data set and the design of the forecasting experiment). In addition, to examine the impact of the length of the estimation period on the inference of the unconditional covariance, I generate forecasts based on SHR-nl using only one year of historical data ($T = 250$) for estimation. I compute estimates of GMVP weights according to the plug-in solution given in Equation (4.4) and the corresponding sequence of portfolio returns, denoted R_{pt}, for a 40-year out-of-sample period spanning 1980 to 2019.

Figure 4.1 illustrates the relative predictive performance over time. I use the demeaned squared portfolio returns as the loss series, where the average loss corresponds to the unconditional variance. The bottom panel shows the cumulative series of pairwise loss differences for DCC-nl and SHR-nl based on $T = 1250$, as well as SHR-nl based on $T = 250$ (SHR-nl*) for the 1980 - 2019 out-of-sample period.[2] In the top panel, the solid lines are moving averages of the standardized 20% loss differences for the out-of-sample period which correspond to test statistics of a test for equal predictive power proposed by Giacomini and Rossi (2010). The so-called fluctuation test rejects the null of equal performance at all time points if the test statistic exceeds the critical value at least once (the dashed black lines represent critical values of 5%). I note that for the DCC-nl vs. SHR-nl comparison, a rejection of a two-sided fluctuation test at the 5% level occurs at the beginning of the out-of-sample period, where the loss of the DCC-nl model is much smaller than that of the SHR-nl model. However, the extreme difference between the performances decreases dramatically when October 19, 1987, known as Black Monday, falls outside the moving average window. As can be seen in the figure below, on this day there is an extreme shift in the cumulative loss series in favor of the DCC-nl model compared to the static models. Presumably, the conditional dynamic approach was more able to capture the emerging volatility in the market just before the Black Monday stock market crash. At the onset of the financial crisis in 2008, in turn, the static models improve over the dynamic DCC-nl. The null of equal predictive performance at all points in time cannot be rejected at any conventional level for DCC-nl vs. SHR-nl*, but SHR-nl* tends to outperform the DCC-nl model on average, as indicated by a negative value of the cumulative loss difference (red solid line in the bottom panel) at the end of the sample. In fact, SHR-nl* appears to have consistently outperformed DCC-nl after October 19, 1987, as indicated by an overall negative trend in the cumulative loss series and negative values of

---

[2]Although not the focus of the analysis, for completeness I also analyzed the results of a DCC-nl based on $T = 250$ observations (DCC-nl*). Since fluctuation test pairwise comparison of DCC-nl and DCC-nl* has revealed that they do not perform significantly different at conventional levels, I omit results for DCC-nl* from the analysis.
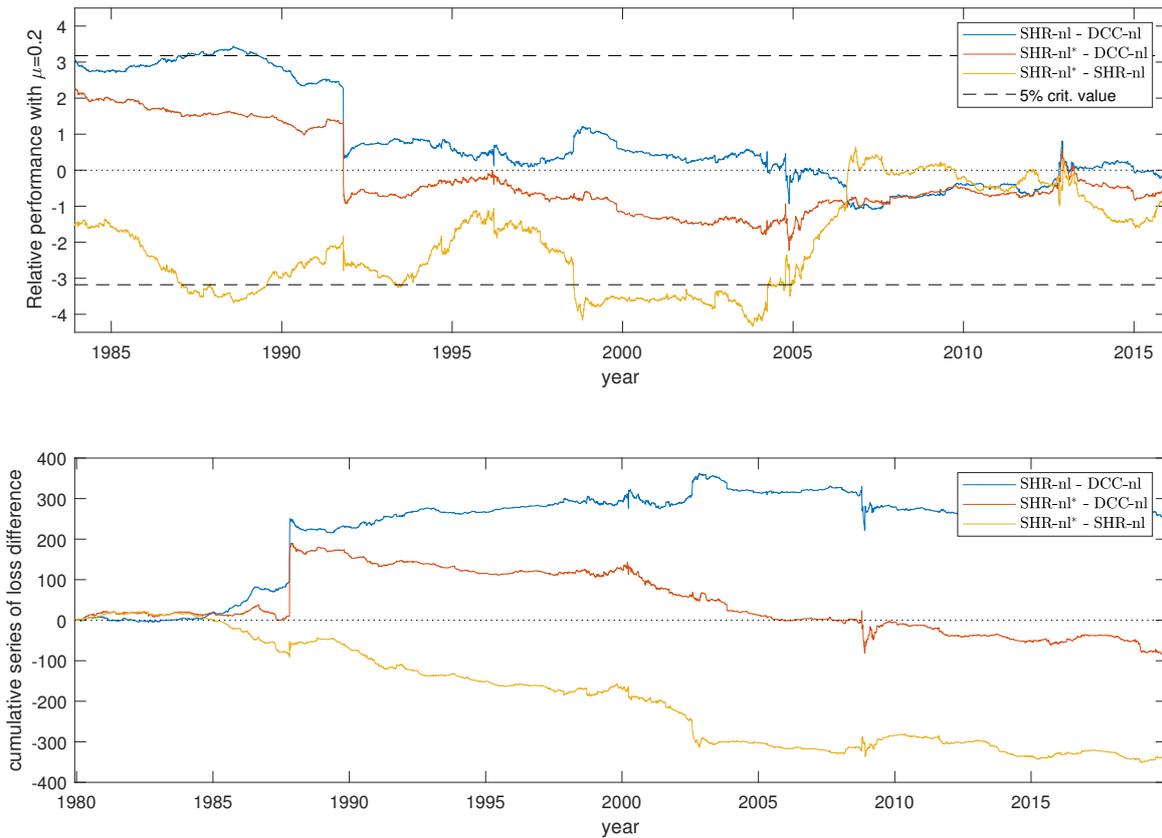
Figure 4.1.: Pairwise comparison of DCC-nl, SHR-nl and SHR-nl* with loss $L_t := (R_{pt} - \bar{R}_{pt})^2$. Top: two-sided GR fluctuation test statistics, black dashed line indicating 5% critical value (Giacomini and Rossi, 2010). On the x-axis, the midpoint of the respective time period to compute the current value of the test statistic is displayed. Bottom: cumulative loss differences.

the fluctuation test statistic. Comparing the two static models, the null of the fluctuation test is clearly rejected, with the model with the shorter estimation window showing better performance. Nevertheless, the fluctuation test statistic exhibits pronounced instabilities in its slope, suggesting that relative predictive performance is not constant over time.

The lesson from this experiment is that the length of the estimation window, i.e., the underlying amount of information, should be carefully chosen when estimating an 'unconditional' covariance matrix and that, taking this into account, a general superiority of the dynamic models cannot be assumed. The lack of temporal stability in relative predictive capabilities suggests that an optimal forecasting model would have to be able to dynamically adopt the characteristics of one or the other approach over time. Moreover, it should be noted that it is difficult to determine which model is best suited in which market situation, as shown, for example, by the different relative performances around Black Monday compared to the beginning of the financial crisis in 2008. From a theoretical perspective, one could argue that in volatile periods, the conditional approaches and models with shorter data histories are able to respond more quickly to changing economic conditions. In turn, the static approaches tend to be more parsimonious, and a longer data history is associated

with a lower concentration rate, so they can be expected to be more robust to estimation noise. These findings motivate the development of my proposed model, which is described in the following sections. The goal is to develop a predictive approach that incorporates the following:

(i) *Accommodating the advantages of flexibility of DCC-nl and robustness of SHR-nl.* To this end, I improve upon unconditional covariance estimation with a shortened estimation window by incorporating more information about remote data, but without giving too much weight to the remote observations. Instead, following the proposal of Bodnar et al. (2021), only deviations from previous timestamps are penalized.

(ii) *Local adaptation to intensity and type of regularization.* As an alternative to shrinkage, I employ a more flexible but equally parsimonious regularization that depends on a small set of penalties that are recalibrated each month.

(iii) *Data-driven and problem-oriented heuristics for this adaption.* Calibration is based on the most recent out-of-sample performance with respect to the evaluation criterion of interest.

4.3. Sparse precision modeling

Among the efficient portfolios defined in Equations (4.1) – (4.3), constructed for the returns on a set of risky assets, the best known selection strategies are the variance-minimizing GMVP and the Sharpe ratio-maximizing tangency portfolio. In this section, I begin by considering the analytical solutions to these allocations in terms of the inverse covariance matrix to motivate my precision modeling approach in general, and then illustrate the augmented time-varying LASSO as an extension of the Gaussian graphical LASSO for precision matrices.

Let $\Theta = \Sigma^{-1}$ be the unconditional precision of R_t. The weight vectors for the global minimum variance portfolio and for the maximum Sharpe ratio portfolio obtain as a scaled linear function of Θ:

$$w^{GMV} = \frac{\Theta\iota}{\iota'\Theta\iota}, \quad w^{maxSR} = \frac{\Theta\mu}{\iota'\Theta\mu}.$$

Most strikingly, the GMVP mapping corresponds to the row or column sums of Θ, scaled by the sum over all its elements. The scaled linear linkage makes Θ much easier to interpret compared to a covariance in the context of portfolio selection, or potentially subject to limitations if the portfolio weights are to be regularized. For example, if one of the row sums is negative, negative GMVP weights occur, which requires that the absolute sum of the weights be greater than one (increased L_1 norm of the weight vector). This holds true because the term $\iota'\Theta\iota$ is always positive for positive definite Θ. Similarly, a large variation in

the weights, i.e., a large deviation from the equally weighted portfolio, is represented by an increased L_2-norm of the weight vector, which corresponds to a large variance in the column sums of Θ. In this paper, I denote by $|| \cdot ||_1$ and $|| \cdot ||_2$ element-wise L_1 and L_2 norms for both vectors and matrices.

In high-dimensional applications, it is usually necessary to impose constraints to reduce estimation noise, for example in the form of sparsity. However, even with hundreds of assets, it is very unlikely that two return series will have zero correlation, especially in highly interconnected financial systems that are typically analyzed in this context. In contrast, as shown in the next subsection, zero entries in the precision matrix in a Gaussian model for returns require conditional independence. Although this assumption may be more realistic, for example, factor models commonly used for financial asset covariances and derived from economic theory (e.g., capital asset pricing model developed by Sharpe, 1964 and Lintner, 1965; three-factor model of Fama and French, 1993) also rule out sparsity in the precision matrix. Therefore, I develop a new penalty function that supports the regularization of portfolio weights but removes the assumption of sparsity induced by the classical LASSO L_1-type regularization.

4.3.1. Gaussian graphical model for precision matrix estimation

In a Gaussian graphical model based on the assumption that the data is multivariate normal distributed, the precision matrix Θ is visualized by an undirected graph. As the normal distribution is fully specified by its first two moments, the potentials, i.e., the values of the edges of the graph corresponding to the entries of the precision matrix, capture all relevant information regarding the relation between the variables. The partial correlations, i.e., the correlation between any two elements $i, j \in (1, \ldots, n), i \neq j$ of the return vector, conditional on all other returns indexed by $-(i, j)$, can be expressed as a function of the elements of the precision:

$$\rho_{i,j|-(i,j)} = \frac{-\Theta_{[ij]}}{\sqrt{\Theta_{[ii]}\Theta_{[jj]}}},$$

where the subscript $[ij]$ denotes the entry in row i and column j of a matrix. Particularly, if an entry $\Theta_{[ij]} = 0$, this implies conditional independence between this pair of returns, which serves as an intuitive interpretation of a sparse graph (details on the derivation see Appendix C.1).

For a sample of return observations $R_t, \ t = 1, \ldots, T$ independent across time, with each $R_t \sim N(\mu, \Theta^{-1})$, one can replace μ by its Maximum Likelihood estimate $\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t$ and write the likelihood function in terms of Θ:

$$L(\Theta) = \frac{1}{(2\pi)^{T/2}} \det(\Theta)^{T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t - \bar{R})'\Theta(R_t - \bar{R})\right).$$

This implies for the log-likelihood

$$\log L(\Theta) \propto T/2 \left(\log(\det(\Theta)) - \text{trace}(S\Theta) \right), \quad S = \frac{1}{T} \sum_{t=1}^{T} (R_t - \bar{R}_t)(R_t - \bar{R}_t)',$$

which is clearly maximized at $\hat{\Theta}_{ML} = S^{-1}$, the inverse of the Maximum Likelihood estimate of the covariance of the data. The sparsity enforcing LASSO regularization of the precision matrix for high-dimensional settings is to add an element-wise L_1-type penalty which implies that some edges of Θ will be zero, which leads to the following solution $\hat{\Theta}^{\text{LASSO}}$:

$$\hat{\Theta}^{\text{LASSO}} = \underset{\Theta \in S_n^{++}}{\text{argmin}} \, T/2 \left(\text{trace}(S\Theta) - \log(\det(\Theta)) \right) + \lambda \sum_{i,j} |\Theta_{i,j}|, \tag{4.6}$$

where $S_n^{++} \subset \mathbb{R}^{n \times n}$ denotes the subspace of positive definite matrices and $\lambda \geq 0$ is a penalty parameter steering the intensity of regularization. Noteworthy, $\hat{\Theta}^{\text{LASSO}}$ also exists for situations with $n > T$ in which the empirical covariance S has reduced rank. Several approaches have been proposed to solve this optimization problem, addressing the computational difficulties of maximization of the not continuously differentiable penalized log-likelihood. For example, Meinshausen et al. (2006) proposed a decomposition into node-wise linear regression problems to save computation time. This approach was recently adapted by Callot et al. (2019) in the context of high-dimensional financial precision matrix estimation. What is widely known as *graphical LASSO* is the algorithm of Friedman et al. (2008), which is a pathwise coordinate descent procedure that modifies the node-wise regression to yield fast convergence to the exact minimizer of the negative of the penalized log-likelihood. The basis of this algorithm and similar approaches is the work of Banerjee et al. (2008), who has shown that Equation (4.6) is indeed a convex optimization problem, which is also the essential condition for determining the tractability of a more general class of penalty functions.

4.3.2. Augmented time-varying graphical LASSO

Although it is widely known that correlation structures between financial assets are typically not stable over time, even recent graphical approaches to estimating the precision matrix for the purpose of portfolio optimization are limited to a static setting (see, e.g., Janková and van de Geer, 2018; Callot et al., 2019; Lee and Seregina, 2021). While it appears to be a straightforward extension of graphical LASSO to account for potential heteroscedasticity in the data, what degree of temporal stability is helpful or harmful for prediction remains an empirical question (see Section 4.2.2). The starting point for my formulation of a time-varying network for Θ is the time-varying Graphical LASSO (TVGL) approach of Hallac et al. (2017), in which an *sliced* interpretation of time is considered. In particular, the joint distribution of returns is assumed to remain stable for a given number of periods. In the terminology of Dahlhaus (1996), the unrestricted model thus implies a locally stationary

process with i.i.d. increments R_t within the slices. The temporal evolution, i.e., changes in neighboring timestamps, can be regularized by penalty expressions in the same way as the slice-individual estimates of the graph itself. The approach is very general and can be applied in many disciplines to a variety of problems (e.g., Hallac et al., 2017 presents an application to automobile sensors). I consider it particularly appropriate for the area of interest examined in this paper, as it is designed to address several important challenges in high-dimensional portfolio allocation problems caused by the bias-variance tradeoff in estimating on the base of noisy financial return observations. The TVGL matches empirical observations by taking into account both sparsity and temporal consistency following the idea that neighboring timestamps should have very similar estimates of the network in most cases, leading to a fairly moderate re-balancing. Moreover, the penalties on the dynamics also implicitly control the degree of temporal stability and the weighting of historical data information, which proved to be very influential for the predictions in the empirical investigation in Section 4.2.2. For out-of-sample predictions, only the precision estimate for the last timestamp is relevant. While this estimate with high values of the penalty parameters for the dynamics strongly incorporates the precision estimates of the previous slices, with zero temporal regulation it is based only on the data from the last slice.

For inferring a time-varying sequence of networks, I set up a sequence of graphical LASSO problems for the precision which are coupled together in a chain to potentially penalize deviations in the estimations. I aim to infer Θ_τ, $\tau = 1, \ldots, T_\tau$, the inverse covariance matrices which are assumed to remain stable for $\mathcal{N}_\tau$ periods each. Precisely, I impose for observations $t = 1, \ldots, T$, with $T = \sum_{j=1}^{T_\tau} \mathcal{N}_j$, that slice τ covers the periods $I_\tau = \{\sum_{j=1}^{\tau-1} \mathcal{N}_j + 1, \sum_{j=1}^{\tau-1} \mathcal{N}_j + 2, \ldots, \sum_{j=1}^{\tau} \mathcal{N}_j\}$ such that the precision at time t is equal to Θ_τ for $t \in I_\tau$, and denote by S_τ the corresponding empirical covariance in this period with sample mean $\bar{R}_\tau$. Conditional on $\bar{R}_\tau$, the Gaussian log-likelihood for the full sample is, up to an additive constant, given as

$$\sum_{\tau=1}^{T_\tau} l_\tau(\Theta_\tau), \quad l_\tau(\Theta_\tau) = \mathcal{N}_\tau(\log(\det(\Theta_\tau)) - \text{trace}(S_\tau \Theta_\tau)).$$

The penalization for the graph itself and its evolution are governed by a set of $J + M$ convex penalty functions $\Psi_j(\cdot)$ and $\tilde{\Psi}_m(\cdot)$ weighted with nonnegative parameters λ_j and β_m, respectively. This leads to the following optimization problem for a given set of parameters $\lambda_1, \ldots, \lambda_J, \beta_1, \ldots, \beta_M$, which I define as augmented time-varying graphical LASSO (ATVGL):

$$\{\hat{\Theta}\}_{\tau=1}^{T_\tau} = \underset{\{\Theta\}_{\tau=1}^{T_\tau} \in S_n^{++}}{\text{argmin}} \sum_{\tau=1}^{T_\tau} \left(-l_\tau(\Theta_\tau) + \sum_{j=1}^{J} \lambda_j \Psi_j(\Theta_\tau) \right) + \sum_{t=2}^{T_\tau} \left(\sum_{m=1}^{M} \beta_m \tilde{\Psi}_m(\Theta_\tau - \Theta_{\tau-1}) \right).$$

(4.7)

In principle this flexible structure allows to enforce virtually any structure for the sequence of precision matrices. The challenging and, to some extent, limiting part here is to be able to solve the problem, i.e., ensuring unique solutions and setting up some algorithm that is ensured to converge to the global minimum (in finite time).

4.3.3. ADMM solution

To this end, I again follow Hallac et al. (2017) and use the alternating direction method of multipliers (ADMM) (Boyd et al., 2011), an iterative, Lagrange-type optimization scheme, and apply it to the generalized problem stated in Equation (4.7). The idea here is to split up the problem along the timestamps and the penalties, into a series of subproblems for which analytical solutions are available. Particularly, for the ADMM the problem is re-formulated such that it consists of two blocks of primal variables, namely, $\bar{\Theta} = (\Theta_1, \ldots, \Theta_{T_\tau})$ and the so called consensus variable $Z = \{Z_0, Z_1, Z_2\} = \{(Z_{1,0,1}, \ldots, Z_{T_\tau,0,J}), (Z_{1,1,1}, \ldots, Z_{T_\tau-1,1,M}), (Z_{2,2,1}, \ldots, Z_{T_\tau,2,M})\}$, where elements of Z_0 are inserted into the penalty functions Ψ_j and elements of the pairs in Z_1, Z_2 into the penalties $\tilde{\Psi}_m$:

$$
\{\hat{\Theta}\}_{\tau=1}^{T_\tau} = \underset{\{\Theta\}_{\tau=1}^{T_\tau} \in S_n^{++}}{\operatorname{argmin}} \sum_{\tau=1}^{T_\tau} \left(-l_\tau(\Theta_\tau) + \sum_{j=1}^{J} \lambda_j \Psi_j(Z_{\tau,0,j}) \right)
$$
$$
+ \sum_{t=2}^{T_\tau} \left(\sum_{m=1}^{M} \beta_m \tilde{\Psi}_m(Z_{\tau,2,m} - Z_{\tau-1,1,m}) \right),
$$

$$
\text{s.t. } Z_{\tau,0,j} = \Theta_\tau \; \forall j, \tau = 1, \ldots, T_\tau; \quad (Z_{\tau-1,1,j}, Z_{\tau,2,j}) = (\Theta_\tau, \Theta_{\tau-1}) \; \forall j, \tau = 2, \ldots, T_\tau.
$$
(4.9)

Incorporating the constraints of Equation (4.9) into Equation (4.8) leads to the unconstrained optimization of the augmented Lagrangian function

$$
\mathcal{L}_\rho(\bar{\Theta}, Z, U) = \sum_{\tau=1}^{T_\tau} \left(-l_\tau(\Theta_\tau) + \sum_{j=1}^{J} \lambda_j \Psi_j(Z_{\tau,0,j}) \right) + \sum_{t=2}^{T_\tau} \left(\sum_{m=1}^{M} \beta_m \tilde{\Psi}_m(Z_{\tau,2,m} - Z_{\tau-1,1,m}) \right)
$$
$$
+ (\rho/2) \sum_j \left(\sum_{\tau=1}^{T_\tau} \left(||\Theta_\tau - Z_{\tau,0,j} + U_{\tau,0,j}||_2^2 + ||U_{\tau,0,j}||_2^2 \right) \right)
$$
$$
+ (\rho/2) \sum_m \left(\sum_{\tau=2}^{T_\tau} \left(||\Theta_{\tau-1} - Z_{\tau-1,1,m} + U_{\tau-1,1,m}||_2^2 + ||U_{\tau-1,1,m}||_2^2 \right.\right.
$$
$$
\left.\left. + \; ||\Theta_\tau - Z_{\tau,2,m} + U_{\tau,2,m}||_2^2 + ||U_{\tau,2,m}||_2^2 \right) \right),
$$
(4.10)

where $U = \{U_0, U_1, U_2\} = \{(U_{1,0,1}, \ldots, U_{T_\tau,0,J}), (U_{1,1,1}, \ldots, U_{T_\tau-1,1,M}), (U_{2,2,1}, \ldots, U_{T_\tau,2,M})\}$ is the scaled dual variable and $\rho > 0$ is the augmented Lagrangian parameter referred to as ADMM penalty parameter. The optimization consists of the following updates, where

k denotes the iteration number and the superscript denotes in which period the respective matrices have been updated:

(a) $\bar{\Theta}^{k+1} := \underset{\{\Theta\}_{\tau=1}^{T_\tau} \in S_n^{++}}{\text{argmin}} \mathcal{L}_\rho(\bar{\Theta}, Z^k, U^k)$

(b) $Z^{k+1} := \begin{pmatrix} Z_0^{k+1} \\ Z_1^{k+1} \\ Z_2^{k+1} \end{pmatrix} := \underset{Z_0, Z_1, Z_2}{\text{argmin}} \mathcal{L}_\rho(\bar{\Theta}^{k+1}, Z, U^k)$

(c) $U^{k+1} := \begin{pmatrix} U_0^{k+1} \\ U_1^{k+1} \\ U_2^{k+1} \end{pmatrix} = \begin{pmatrix} U_0^k \\ U_1^k \\ U_2^k \end{pmatrix} + \begin{pmatrix} (\underbrace{\Theta_{\{1:T_\tau\}}^{k+1}, \ldots, \Theta_{\{1:T_\tau\}}^{k+1}}_{\#J}) - Z_0^{k+1} \\ (\underbrace{\Theta_{\{1:T_\tau-1\}}^{k+1}, \ldots, \Theta_{\{1:T_\tau-1\}}^{k+1}}_{\#M}) - Z_1^{k+1} \\ (\underbrace{\Theta_{\{2:T_\tau\}}^{k+1}, \ldots, \Theta_{\{2:T_\tau\}}^{k+1}}_{\#M}) - Z_2^{k+1} \end{pmatrix}.$

I initialize Z^k and U^k by appropriately stacked identity matrices of dimension n. Notably, all solutions can be obtained separately for each Θ_τ, $Z_{\tau,0,j}$ and $(Z_{\tau-1,1,m}, Z_{\tau,2,m})$. I specifically make use of several proximal operators (Boyd et al., 2011), which define a class of particular optimization problems for which well-known closed form solutions exist. A proximal operator considers a trade-off between minimizing some convex function f with respect to a variable X and minimizing the squared Euclidean distance of X to a target A (details are provided in Appendix C.2). For solving (a), I apply a proximal operator for the negative log-likelihood and the components in the second to fourth line of Equation (4.10) which depend on Θ_τ (see Appendix C.3.1). The solution in (b) depends on the functional form of the particular penalties Ψ_j and $\tilde{\Psi}_m$ and is described for the set of employed penalties in Section 4.3.4.

The two-block ADMM utilized here is guaranteed to converge to the global optimum as long as the penalty functions Ψ_m and $\tilde{\Psi}_j$ are convex in Z, given that the negative log-likelihood is clearly strictly convex in $\{\Theta\}_{\tau=1}^{T_\tau}$. The algorithm uses a stopping criterion based on the primal and dual residuals being below specified thresholds (Boyd et al., 2011) which are set to very conservative values in my implementation (Matlab 2021a).

4.3.4. Penalty functions

The precision matrix estimates obtain as optimum in a trade-off between fit and sparsity, both with respect to the precision at each slice τ as well as its degree of time-variation. The former is accounted for by the terms $\lambda_j \Psi_j(\cdot)$ while the latter is accounted for by the terms $\beta_m \tilde{\Psi}_m(\cdot)$. Although in principle all convex functionals are possible, in practice, the computing time is the bottleneck and needs to be carefully accounted for. As all subproblems need to be solved for T_τ or $T_\tau - 1$ times in every iteration, only functions with closed form solutions can be used in practice and the number of penalties should be kept to a low number

in order to alleviate convergence. Furthermore, the computational costs depend linearly on the number of timestamps.

I set $M = 1$, imposing a simple element-wise L_1 penalty, $\tilde{\Psi}_1(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} |X_{[ij]}| = ||X||_1$, inserting for X the difference $Z_{r-1,1,m} - Z_{r,2,m}$ to limit variation in subsequent timestamps. Initial experiments showed that L_1 and L_2 entrywise penalization lead to qualitatively similar results in my application. The *a few edges at a time penalty*, as phrased by Hallac et al. (2017), can be solved by applying a soft-thresholding proximal operator that particularly exploits the element-wise structure of $\tilde{\Psi}_1$ (see Appendix C.3.2).

Regarding the estimation of the precision of the individual slices τ, I impose $J = 2$ penalties, which facilitates to relate the specification to the application of interest, i.e., portfolio allocation. The first is again a standard element-wise LASSO penalty, in the terminology of Hallac et al. (2017) the *a few edges at a time omitting the diagonals* penalty $\Psi_1(X) = \sum_{i,i \neq j}^{n} \sum_{j=1}^{n} |X_{[ij]}|$, inserting for X the matrices $Z_{\tau,0,1}$. The diagonal elements of the estimates for Θ_τ are not addressed here because it is not sensible to assume sparsity for them: For $\lambda_1 = \infty$, the penalty Ψ_1 leads to a diagonal precision with all remaining edges equal to zero, which corresponds to a diagonal covariance matrix leaving the variances unaffected. The soft-thresholding proximal operator can directly be applied to solve the problem (see Appendix C.3.2). With the second regularization, I intend to relate the problem to my application of interest and propose a new penalty function that aims to lower economically implausible allocations. Although conceptually not restricted to that, I focus here on the minimum variance allocation that is exclusively determined by the covariation among the data and avoids estimation of the mean return which is known to be hard to estimate from return series at low sampling frequency like daily data with a typically low signal-to-noise ratio. A penalty function of which the functional form or the shrinkage target is highly dependent on a noisy estimation of the mean return cannot be assumed to work well in out-of-sample predictions. Moreover, I desire a penalty which complements Ψ_1 in reducing estimation noise, but without demanding sparsity in Θ_τ, motivated by the fact that, e.g., with the presence of an underlying common market factor, the assumption of conditional independence can be too rigid. As I particularly want to limit extreme portfolio weights on particular stocks, I consider a squared euclidian distance of the GMVP weights from the equally weighted portfolio which translates to L_2-type regularization. Implicitly, this also limits gross exposure and leads for infinite penalization to the equally weighted portfolio. For deriving such a penalty, my starting point are the GMVP weights in terms of the precision matrix, i.e., $w^{GMV} = \frac{\Theta' \iota}{\iota' \Theta \iota}$. Conceptually, I could directly apply an L_2 norm constraint on these weights, or equivalently, on their deviations from the equally weighed portfolio ι/n. The problem is that the metrics inducing L_p norms do not lead to a closed form solution for Θ. To illustrate, note that for some matrix X representing a precision,

$$||w^{GMV} - \iota/n||_2^2 = ||\frac{1}{\iota' X \iota}(X\iota - \iota' X\iota/n)||_2^2,$$

where, due to the factor $(\iota' X \iota)^{-1}$, the first order condition is highly nonlinear in X, an analytical solution would be prohibited and computationally more costly numerical optimization techniques would be required. Instead, I consider a scaled version, namely,

$$\Psi_2(X) = ||(X\iota - \iota' X \iota/n)||_2^2,$$

which is a quadratic convex problem with linear first order conditions, minimized for all symmetric generalized doubly stochastic matrices[3], which imply an equally weighted GMVP (for X, the matrices $Z_{\tau,0,2}$ are inserted). Particularly, this includes estimates for the precision which imply an equicorrelation covariance matrix for the joint return distribution. An intuitive interpretation of the newly introduced penalty Ψ_2 is that it is proportional to the variance in the row and column sums of X:

$$\Psi_2(X) = ||(X\iota - \iota' X \iota/n)||_2^2$$
$$= \sum_i \left(\left(\sum_j X_{[i,j]} \right)^2 - 2 \sum_j X_{[i,j]} \left(\sum_{i,j} X_{[i,j]}/n \right) + \left(\sum_{i,j} X_{[i,j]}/n \right)^2 \right)$$
$$= \sum_i \left(\sum_j X_{[i,j]} \right)^2 - n \left(\sum_{i,j} X_{[i,j]}/n \right)^2,$$

which is clearly minimized if all row and column sums are equal, but with no direct restriction on the sum of the entries itself. It is therefore optimally suited to augment the LASSO penalty Ψ_1. For some component $Z_{\tau,0,2}$, the vectorized (vec) solution under parameters λ_2 and ρ is given by

$$\underset{\text{vec}(Z_{\tau,0,2})}{\operatorname{argmin}} \mathcal{L}_\rho(\Theta_\tau, Z_{\tau,0,2}, U_{\tau,0,2}) = \underset{\text{vec}(Z_{\tau,0,2})}{\operatorname{argmin}} \lambda_2 \Psi_2(Z_{\tau,0,2}) + \rho/2 || (Z_{\tau,0,2} - (\Theta_\tau + U_{\tau,0,2})) ||_2^2$$
$$= \left(\underbrace{\lambda_2 [(1_{n \times n} \otimes I_n) - (I_n \otimes 1_{n \times n}) - 2/n 1_{n^2 \times n^2}] + \rho I_{n^2}}_{:= J_{\Psi_2}} \right)^{-1} \rho \text{vec}(\Theta_\tau + U_{\tau,0,2}),$$

whereby the inversion of the $n^2 \times n^2$ matrix J_{Ψ_2} only needs to be performed once. In Appendix C.3.3, I derive the analytical solution for $J_{\Psi_2}^{-1}$ by exploiting the particular structure of J_{Ψ_2} and show that it only consists of three distinct entries as functionals of ρ, λ_2 and n. Therefore, I can circumvent the vast-dimensional matrix multiplication that has to be conducted to solve for all $Z_{\tau,0,2}$ in each iteration k and replace it by computing appropriately weighted sums of the elements of $\rho \text{vec}(\Theta_\tau + U_{\tau,0,2})$. This reduces computing time by more than 98% and makes the penalty function Ψ_2 scalable to large and vast dimensional applications.

Figure 4.2 illustrates the effects of the different penalties that are being used in my ATVGL

---

[3] A doubly stochastic matrix (DSM) is a square matrix of nonnegative real numbers, each of whose rows and columns sums to one. For generalized DSM, the nonnegativity condition is dropped.
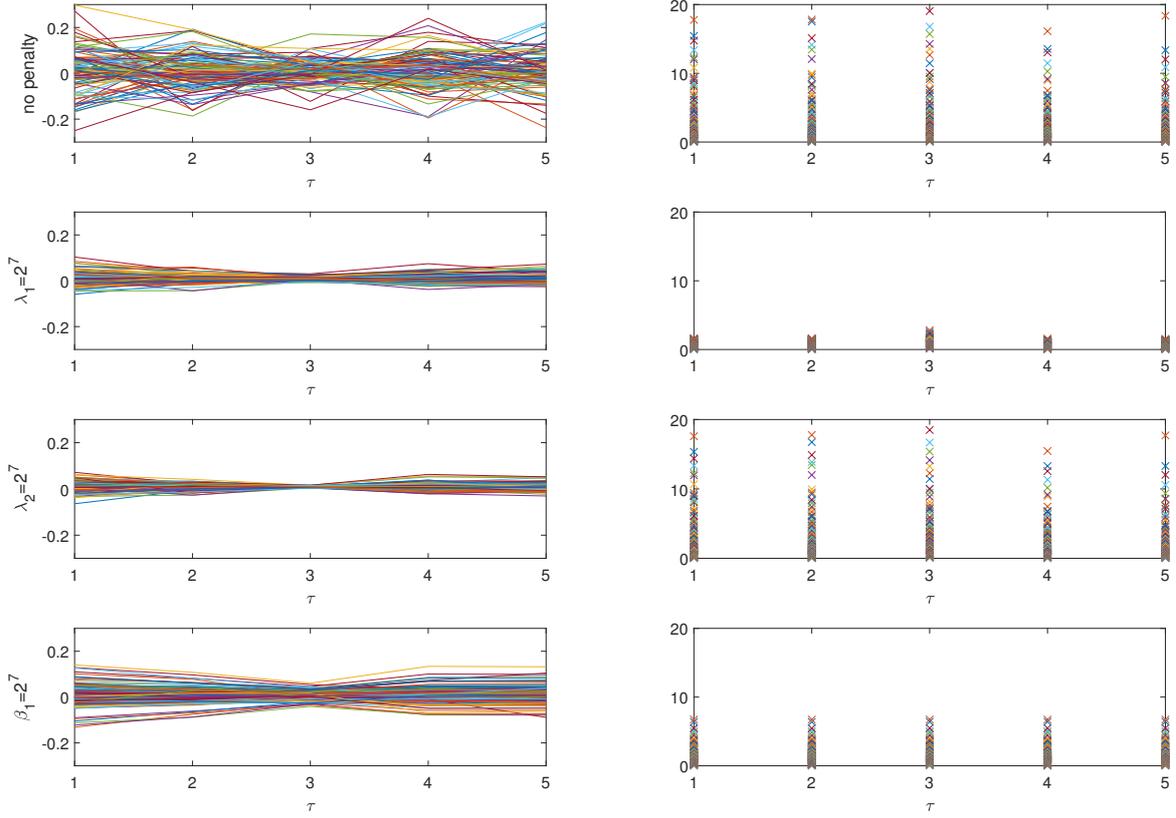
Figure 4.2.: In-sample comparison of penalty functions with slice length $\mathcal{N}_\tau = 250$, $n = 100$, based on time period 2015 - 2019 ($T = 1250$). Left: GMVP weights for slices $\tau = 1, \ldots, 5$, right: eigenvalues of Θ_τ, $\tau = 1, \ldots, 5$. From top to bottom: $(\lambda_1, \lambda_2, \beta_1) = (0, 0, 0)$, $(\lambda_1, \lambda_2, \beta_1) = (2^7, 0, 0)$, $(\lambda_1, \lambda_2, \beta_1) = (0, 2^7, 0)$, $(\lambda_1, \lambda_2, \beta_1) = (0, 0, 2^7)$.

specification, that is, $\lambda_1 \Psi_1(\cdot)$, $\lambda_2 \Psi_2(\cdot)$ and $\beta_1 \tilde{\Psi}_1(\cdot)$. For return series of $n = 100$ assets from the data set used in my empirical application, the penalized log-likelihood is maximized for a sample of size $T = 1250$ in the period $2015 - 2019$, divided in slices each of length $\mathcal{N}_\tau = 250$, under different constellations of the of the penalty parameters: an unrestricted model with $\lambda_1 = \lambda_2 = \beta_1 = 0$, as well as a model with each one of λ_1, λ_2 and β_1 set to 2^7 with the remaining parameters equal to zero. The left panel displays the resulting GMVP weight estimates and the right panel presents eigenvalues of the estimates for $\{\Theta\}_{\tau=1}^{T_\tau}$ on each timestamp. In contrast to the unconstrained model, where $\{\hat{\Theta}\}_{\tau=1}^{T_\tau}$ is identical to the inverse of the empirical covariance of each slice, visual inspection shows that each of the three penalties is able to regularize the portfolio weights. It is shown that taking into account information from neighboring timestamps over $\tilde{\Psi}_1$ leads to an 'averaging' of the precision estimates and hence the allocations, which also leads to a significant decrease in the dispersion of the eigenvalues of each $\hat{\Theta}_\tau$. Similarly, one can see that both Ψ_1 and Ψ_2 imply a shrinkage of portfolio weights towards the equally weighted portfolio. However, under the standard element-wise LASSO-type penalty, the reduction is associated with an extreme shrinkage of the eigenvalues of each $\hat{\Theta}_\tau$. The average L_1 norm of the precision estimates

(not reported here) is much smaller than under the two other regularizations implying that it enforces sparsity in the network. In contrast, under Ψ_2, the eigenvalues of $\{\hat{\Theta}\}_{\tau=1}^{T_\tau}$ are hardly affected, as the regularization of the weights is only driven by more homogeneity in the row and column sums. Overall, this figure shows that the three penalties leads to a regularization of the portfolio weights associated with very different effects on the precision estimates. While Ψ_1 and Ψ_2 affect the structure of the networks on each slice, $\tilde{\Psi}_1$ controls the homogeneity along the slices and thus the amount of remote data information considered for estimation.

4.4. Empirical application

4.4.1. Data and design of the experiment and benchmark models

I use the data set analyzed by Moura et al. (2020) which consists of the daily prices of all NYSE, AMEX and NASDAQ stocks and update it to include price observations until the end of the year 2019. The full sample covers the period from 12/28/1973 to 11/26/2019 for a total of $11,582$ trading days. In my out-of-sample experiments I focus on one-day-ahead forecasts obtained by re-estimating the model parameters every month on a rolling window scheme, where I follow the convention that 21 consecutive days constitute one month. The out-of-sample period starts on 12/07/1979 and ends on 11/26/2019 which results in a total of 10,080 point forecasts, corresponding to 480 months or 40 years. I conduct an extensive study of GMVP predictions with portfolio size $n = 100$ (Section 4.4.3), and additionally provide an analysis of mean-variance optimal allocations as well as results for cross-sectional dimension $n = 200$ to robustify my findings (Section 4.4.4). I fit the length of the estimation window to five years ($T = 1250$) additional to one year (T = 252) for my suggested approach due to the design of my parameter calibration scheme (see Section 4.4.2). This leads to an effective information set of six years of past data for my ATVGL model in contrast to five years for the benchmark approaches. However, a six year estimation period leads for none of the benchmark models to an improved performance (not reported here). Therefore I rely on the commonly used five year window width. Following Engle et al. (2019), the n stocks included in a portfolio are re-determined before re-estimating the parameters each (virtual) month. They are selected as follows: First, I identify the stocks that have a complete series of reported returns over the most recent T days and over the next 21 days. Then, I identify all pairs of stocks with a sample correlation larger than 0.95 over the past T days and remove the respective stock with lower trading volume observed at the time of re-estimation. Finally, I select the largest n stocks in terms of market capitalization at the re-estimation period.

As alternatives to my proposed GMVP models I use the following static and dynamic approaches from the literature: (i) The plug-in estimator constructs the prediction of the GMVP weights by estimating the return unconditional covariance matrix Σ with the sample covariance matrix. (ii) The linear shrinkage (SHR-l) estimator modifies the plug-in estima-

tor by estimating Σ via the linear shrinkage approach of Ledoit and Wolf (2004). (iii) The nonlinear shrinkage (SHR-nl) approach estimates Σ via the nonlinear shrinkage procedure of Ledoit and Wolf (2012, 2015). While the linear shrinkage estimator shrinks all sample eigenvalues towards toward the grand mean of the sample eigenvalues with the same intensity, the nonlinear shrinkage approach uses an individualized intensity for each eigenvalue. (iv) The linear shrinkage market factor (SHR-l-1f) estimator modifies the plug-in estimator by estimating Σ via the linear shrinkage approach in direction of a market factor covariance matrix (Ledoit and Wolf, 2003) (v) The naïve estimator sets the prediction of the GMVP weights equal to the weights of the equally weighted (naïve) portfolio. (vi) The Gaussian DCC model of Engle (2002) with correlation targeting based on the standard sample covariance matrix (DCC-s). (vii) The Gaussian DCC model with correlation targeting based on nonlinear shrinkage estimates (DCC-nl) (Engle et al., 2019). (viii) An approximate one-factor model with residual covariance according to a Gaussian DCC model with correlation targeting based on nonlinear shrinkage estimates (AFM-DCC-nl) (De Nard et al., 2021). (iv) A regularized exponential Recursive Least Squares with forgetting factor scheme (RLS-REF) to directly infer the GMVP weights (Reh et al., 2021), based on a linear regression of a benchmark return on the return differences to this benchmark, in which the regression coefficients represent the GMVP weights (Kempf and Memmel, 2006).

4.4.2. Dynamic recalibration

An essential step in training a regularized model is the calibration of the hyper-parameters. I want to determine the optimal trade-off between fit and sparsity taking into account that I am solely interested in out-of-sample performance, that data is in temporal order and that it can be assumed that the optimal calibration is time-variant, such that it is sensible to consider a rolling window scheme like for the estimation window. Moreover, I aim to target the application of interest. In Figure 4.3, I graphically illustrate my calibration and forecasting scheme, that I developed based on these considerations. On a log-scaled grid of $15 \times 15 \times 15$ values for the three penalty parameters, that is, $(\lambda_1, \lambda_2, \beta_1) \in \{2^k, k \in \{-7, 7\}\}$, I re-estimate the model every month, based on the current data set determined as described in Section 4.4.1, and compute one month of out-of-sample returns with weights according to the strategy of interest (like, e.g., the unrestricted GMVP, or some gross-exposure constrained mean-variance optimal portfolio) based on the estimated precision of the last slice in the sample. In the vein of Jegadeesh and Titman (1993), I aim to exploit momentum performance and select in each month the set of parameters which yields the best performance evaluated on the criterion of interest (e.g., unconditional variance) within the most recent 12 months of pseudo-out-of-sample period. That is, I consider each the latest 12 evaluation windows as calibration period. Another degree of freedom is the slice length. For the sake of practicability and computational cost, however, I restrict myself to two (fixed) slice lengths: $\mathcal{N}_\tau = 125$ and $\mathcal{N}_\tau = 250$ $\forall \tau$, corresponding to 10% and 20% of the estimation
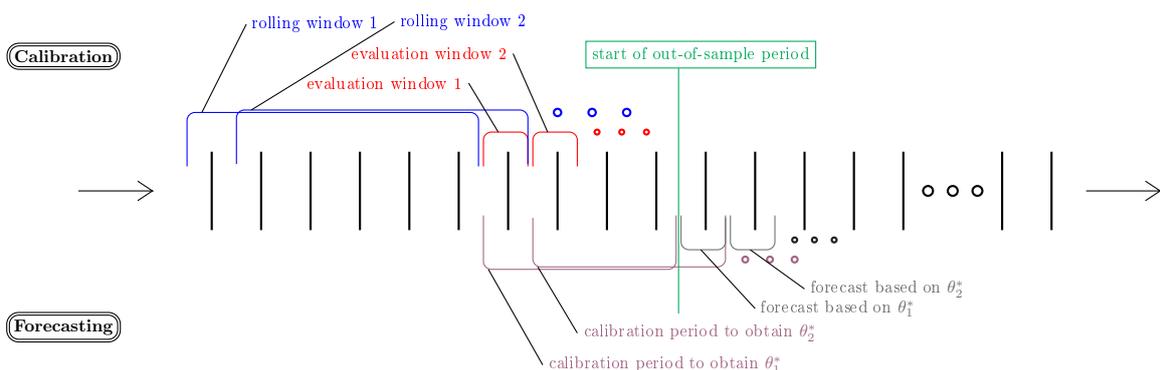
Figure 4.3.: Dynamic calibration scheme for penalty parameters and slice length $\theta = (\lambda_1, \lambda_2, \beta_1, \mathcal{N}_\tau)$. Blue sections: rolling windows for re-estimation ($T = 1250$), red sections: corresponding evaluation windows ($T^E = 21$), violet sections: calibration periods comprising 12 evaluation windows ($T^C = 12T^E = 252$), gray sections: corresponding out-of-sample forecasting periods ($T^F = 21$).

window. Imposing time-stability for 125 or more trading days does not allow for precise inference in the context of event detection for which the TVGL was originally invented, but in order to obtain robust estimates for the inverse of the joint covariance matrix of noisy daily return observations, it can be assumed that the use of large timestamps is beneficial. The choice of $\mathcal{N}_\tau = 250$ is motivated by the results in Section 4.2.2 which indicate that even static models can compete with dynamic specifications when cutting their estimation window to one year (for given cross-sectional dimension $n = 100$). For investigating the benefits of increased flexibility, I additionally fit the model for $\mathcal{N}_\tau = 125$. In total, this leads to $2 \times 15^3 = 6,750$ possible models that are re-estimated every month.

4.4.3. Results

In this section I present the forecasting performance for the ATVGL and its competitors based on several variations of the GMVP with cross-sectional dimension $n = 100$. The minimum variance allocation depends solely on the asset covariation and represents my main analysis for evaluating the time-varying LASSO approach to inferring $\{\Theta\}_{\tau=1}^{T_\tau}$. For a given asset (conditional) covariance estimate $\hat{\Sigma}$, $\hat{\Sigma}_{T+1}$ or the respective precision estimate $\hat{\Theta}_{T_\tau}$, the GMVP weight prediction is given by plugging into Equation (4.4). For the RLS-REF approach the GMVP weight estimates obtain directly from the recursion. In large scale applications as considered here, it might occur that covariance predictions are not well conditioned possibly resulting in unrealistically volatile or economically implausible allocations. Therefore I additionally consider short-selling constrained allocations. These additional restrictions are imposed to the minimization problem stated in Equations (4.1) − (4.2) by adding the following norm constraints:

$$||w||_1 \leq 1 + 2s, \tag{4.11}$$
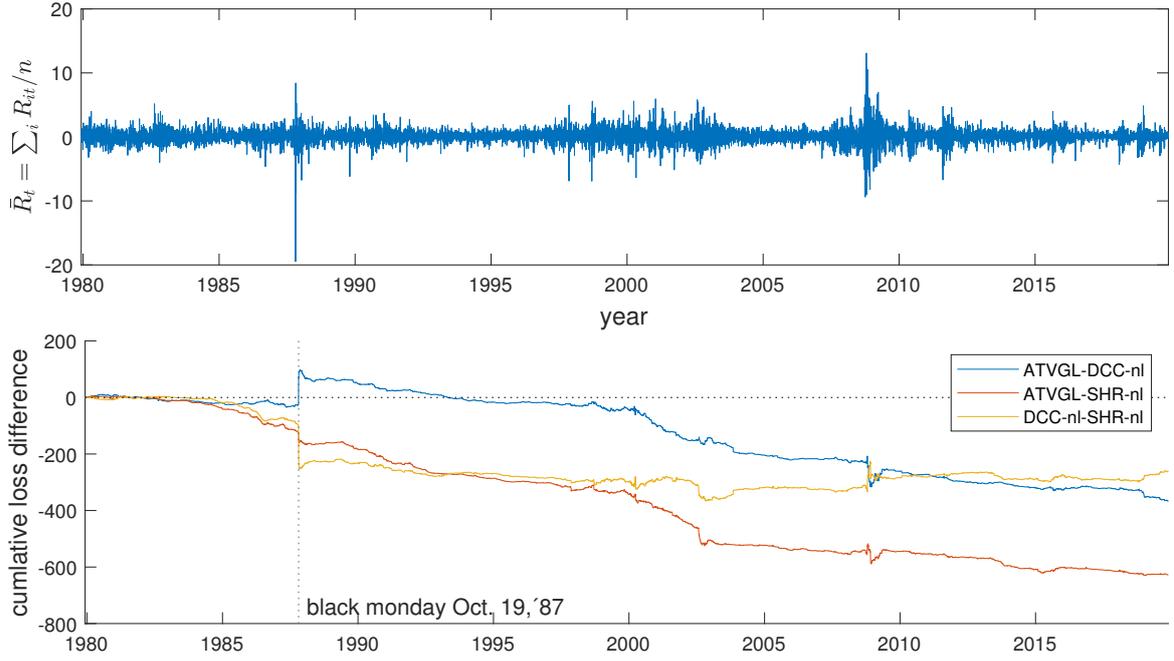
Figure 4.4.: Out-of-sample time plots. Top: daily returns on equally weighted portfolio, bottom: pairwise cumulative loss differences of ATVGL, DCC-nl and SHR-nl with loss $L_t := (R_{pt} - \bar{R}_{pt})^2$.

where $s \geq 0$ denotes the fraction that is allowed to be held short. In the present paper I consider, besides the unrestricted GMVP, the commonly used 130/30-portfolio ($s = 0.3$) as well as a long only 100/0-portfolio ($s = 0.0$). As the constrained minimization problem has no closed-form solution, I rely on the CVX package in Matlab of Grant and Boyd (2014) to solve for the constrained GMVP weights. The optimal calibrations of $\theta = (\lambda_1, \lambda_2, \beta_1, \mathcal{N}_\tau)$ for the ATVGL are selected according to the calibration scheme outlined in Section 4.4.2 each based on the respective allocation rule. I present the empirical out-of-sample portfolio variance $\hat{\sigma}^2$ of ATVGL as well as the dynamic and static benchmark approaches. For some sequence of portfolio returns R_{ps}, $s = 1, \ldots S$, with empirical mean $\bar{R}_p = \frac{1}{S} \sum_{s=1}^{S} R_{ps}$, it is computed as

$$\hat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^{S} (R_{ps} - \bar{R}_p)^2.$$

For assessing the statistical significance of differences in the empirical out-of-sample variance, I apply the model confidence set (MCS) approach of Hansen et al. (2011), taking the demeaned squared returns as loss series.

The top panel in Figure 4.4 plots the times series of the daily returns on the equally weighted portfolio comprising all $n = 100$ assets over the complete out-of-sample period. The bottom panel displays pairwise cumulative loss differences of ATVGL, DCC-nl and SHR-nl. It reveals that except October 19, 1987, in which the DCC-nl allocation leads to a substantially less negative return compared to the two other selection strategies, resulting in

Table 4.1.: Out-of-sample empirical variance unrestricted GMVP for $n = 100$

model/time	full	'80-'84	'85-'89	'90-'94	'95-'99	'00-'04	'05-'09	'10-'14	'15-'19
ATVGL	0.504	0.389	0.744	0.281	0.548	0.661	0.659	0.321	0.424
DCC-nl	0.541	0.404	0.685	0.338	0.564	0.791	0.705	0.367	0.463
DCC	0.557	0.416	0.719	0.353	0.574	0.816	0.713	0.376	0.475
AFM-DCC-nl	0.553	0.409	0.795	0.376	0.582	0.756	0.690	0.355	0.451
SHR-nl	0.566	0.416	0.857	0.366	0.586	0.810	0.678	0.371	0.441
SHR-nl*	0.533	0.419	0.804	0.298	0.573	0.704	0.689	0.332	0.438
SHR-l	0.581	0.430	0.891	0.382	0.597	0.827	0.689	0.376	0.451
SHR-l-1f	0.568	0.419	0.847	0.379	0.589	0.813	0.675	0.368	0.446
plug-in	0.582	0.430	0.893	0.385	0.597	0.828	0.692	0.377	0.451
RLS-REF	0.529	0.418	0.800	0.326	0.572	0.722	0.644	0.326	0.416
naïve	1.209	0.984	1.433	0.598	1.079	1.643	2.185	1.028	0.710

Note: Grey shaded cells indicate that the model belongs to the 90 % MCS.

a shift in the cumulative loss differentials, the ATVGL virtually uniformly outperforms the DCC-nl model, as indicated by an overall negative trend in the cumulative loss differential (blue line). The strength of the approach, however, is that it also uniformly outperforms the static SHR-nl (red line), whereas the relative performance of DCC-nl and SHR-nl doest not appear to be stable over time (yellow line, see also the discussion in Section 4.2.2). In Table 4.1 I report the out of sample empirical variance of ATVGL and all considered benchmark approaches over the full sample as well as for eight subsamples of five years each. With SHR-nl* I denote again the SHR-nl estimate of the unconditional covariance based on a shortened estimation window of the previous $T^* = 250$ trading days. Over the full sample, ATVGL is the best performing approach and the only one to be included in the 90% MCS. Further, it is the only approach to be included in the 90% MCS in all subsamples demonstrating its robustness and flexibility which leads to this uniformly well performance. Among the benchmark models, the overall lowest variance is generated by RLS-REF and SHR-nl*, which both cover distinct features which set them apart from the remaining approaches: RLS-REF directly models the GMVP weights without detour on the covariance, hence it is explicitly designed to optimize the quantity of interest, the unconditional variance. The SHR-nl* only takes into account the previous year of data for a 'pseudo'-unconditional estimation of the joint covariance which also proves to be beneficial in terms of forecasting. However, the results also indicate that it would be helpful to switch between the estimation window lengths, as in three of the eight subsamples, SHR-nl leads to a lower variance than SHR-nl*.

Noteworthy, the comparably well performance of DCC-nl on average can be traced back to the time period $'85-'89$ including Black Monday, whereas it has no longer been competitive (included in 90% MCS) in the last 20 years of the out-of-sample period.

Table 4.2.: Out-of-sample empirical variance exposure-constrained GMVP for $n = 100$

model/time	full	'80-'84	'85-'89	'90-'94	'95-'99	'00-'04	'05-'09	'10-'14	'15-'19
				130/30 strategy					
ATVGL	0.519	0.398	0.829	0.288	0.542	0.663	0.701	0.317	0.411
DCC-nl	0.535	0.418	0.720	0.325	0.568	0.739	0.699	0.363	0.440
DCC	0.540	0.422	0.740	0.330	0.570	0.751	0.698	0.364	0.440
AFM-DCC-nl	0.557	0.420	0.813	0.359	0.583	0.734	0.727	0.368	0.444
SHR-nl	0.567	0.422	0.856	0.364	0.578	0.794	0.708	0.375	0.437
SHR-nl*	0.525	0.411	0.854	0.296	0.560	0.677	0.669	0.322	0.408
SHR-l	0.572	0.425	0.873	0.370	0.580	0.798	0.711	0.374	0.439
SHR-l-1f	0.569	0.422	0.849	0.381	0.577	0.794	0.709	0.372	0.441
plug-in	0.572	0.425	0.872	0.372	0.580	0.798	0.710	0.374	0.439
naïve	1.209	0.984	1.433	0.598	1.079	1.643	2.185	1.028	0.710
				100/0 strategy					
ATVGL	0.596	0.466	0.929	0.328	0.606	0.732	0.895	0.399	0.412
DCC-nl	0.597	0.492	0.781	0.338	0.655	0.786	0.848	0.419	0.447
DCC	0.599	0.494	0.799	0.340	0.653	0.787	0.849	0.418	0.446
AFM-DCC-nl	0.624	0.499	0.877	0.379	0.643	0.793	0.895	0.431	0.467
SHR-nl	0.636	0.498	0.927	0.394	0.638	0.843	0.892	0.440	0.454
SHR-nl*	0.600	0.472	0.965	0.346	0.629	0.720	0.847	0.398	0.418
SHR-l	0.640	0.502	0.946	0.402	0.638	0.845	0.889	0.440	0.455
SHR-l-1f	0.637	0.500	0.923	0.407	0.637	0.841	0.888	0.438	0.456
plug-in	0.641	0.502	0.946	0.404	0.638	0.845	0.890	0.441	0.455
naïve	1.209	0.984	1.433	0.598	1.079	1.643	2.185	1.028	0.710

Note: Grey shaded cells indicate that the model belongs to the 90 % MCS.

The unconditional portfolio variances for allocations imposing gross exposure constraints are collected in Table 4.2. Overall, the results are qualitatively similar to the unconditional GMVP. On average, the allocation strategies become harder to be distinguished, e.g., in the subsample $'05-09'$, nine of the ten approaches are included in the 90% MCS for both exposure constrained strategies. However, both for 130/30 and 100/0, the ATVGL is still

the approach with the smallest unconditional variance over the full sample and the only one to be included in the 90% MCS for all subsamples except for the period $'85 -'89'$. Notably, the RLS-REF strategy can no longer be applied here as this direct modeling approach of the portfolio weights comes with the disadvantage that it does not allow to impose norm constraints on the weights ex-post.

Table 4.3.: ATVGL performance with restricted calibration

θ restriction	$\hat{\sigma}^2$ rel. to none (in %)	avg. $||\hat{\Theta}||_1$	avg. $||\hat{w}||_2$
none	0.0	0.249	0.338
$\lambda_1 = 0$	+6.6	0.606	0.352
$\lambda_2 = 0$	+1.5	0.222	0.343
$\beta_1 = 0$	+1.5	0.368	0.327
$\mathcal{N}_\tau = 125$ only	+1.2	0.222	0.328
$\mathcal{N}_\tau = 250$ only	+0.5	0.249	0.334

Note: Out-of-sample model comparison under several parameter restrictions on the calibration. Left column: empirical variance relative to unrestricted calibration, middle column; average element-wise L_1 norm of precision estimates, right column: average L_2 norm of unrestricted GMVP weight estimates.

The ATVGL approach is driven by four parameters $\theta = (\lambda_1, \lambda_2, \beta_1, \mathcal{N}_\tau)$, for which I fit the model on a grid of dimension $15^3 \times 2 = 6,750$. In my forecasting experiment, the average calibrations are given as $\bar{\lambda}_1^* = 21.18$ (standard deviation $std = 29.88$), $\bar{\lambda}_2^* = 11.76$ ($std = 31.90$), $\bar{\beta}_1^* = 29.61$ ($std = 62.35$), and in 40.74% (59.26%) slice length $\mathcal{N}_\tau = 250$ ($\mathcal{N}_\tau = 125$) is selected for the unrestricted model. For the gross-exposure constrained models, average values of the penalty parameters and the preference for slice length $\mathcal{N}_\tau = 125$ tends to increase slightly. Notably, a mere analysis of the absolute values of the parameters is not sufficient as it does not account for potentially different sensitivities of the penalized log-likelihood with respect to the parameters. For practical purposes, however, it may be desirable to analyze which of the penalties is most important, and which slice length leads to the best performance. To this end, I show in Table 4.3 for the unconditional GMVP strategy, how it alters the results if I restrict each one of the penalty parameters to zero or allow only for one particular slice length. Restricting λ_1 to zero clearly leads to the largest decrease in performance (+6% unconditional variance compared to the unrestricted calibration setting), whereas setting λ_2 or β_1 to zero only leads to a moderate decrease in performance of each approximately 1.5%. Additional to the impact on conditional variance displayed in the first column, the second and third column of Table 4.3 display the average element-wise L_1 norm of the precision (avg. $||\hat{\Theta}||_1$) and the average L_2 norm of the selected out-of-sample GMVP weights (avg. $||\hat{w}||_2$), respectively: They reveal that without the newly introduced penalty Ψ_2, the average dispersion of the weights is slightly larger (0.343 compared to 0.338), but at the same time, substantially more sparsity in the precision is required (average element-wise

L_1 norm of 0.222 compared to 0.249). This again demonstrates the functioning of the newly introduced penalty Ψ_2 used to augment the LASSO caused by Ψ_1: increased homogeneity in the portfolio weights associated with less enforcement of potentially economically implausible sparsity in the precision. Restricting the slice length to be fixed to $\mathcal{N}_\tau = 250$ only leads to a 0.5% increase in unconditional variance. At the same time, doubled number of slices (10×125 compared to 5×250) doubles the computing time such that the computational costs can be cut by 75% when restricting the slice length to $\mathcal{N}_\tau = 250$. Hence, from a practitioner's point of view, if a restriction shall be imposed to shrink the parameter space, considering only slice length $\mathcal{N}_\tau = 250$ appears to be the most attractive one.

4.4.4. Robustness checks

As a second out-of-sample portfolio allocation exercise I consider a mean-variance portfolio based on an investor who aims at minimizing the portfolio risk subject to a target portfolio return μ^*. The solution to the problem defined in Equations (4.1) – (4.3) obtains as:

$$w^{MV} = \operatorname*{argmin}_{w} w'\Sigma w, \quad \text{s.t. } \iota'w = 1,\ m'w = \mu^* \tag{4.12}$$

$$= \Sigma^{-1}\frac{m(Cb - D) + \iota(E - Db)}{EC - D^2}, \tag{4.13}$$

$$C = \iota'\Sigma^{-1}\iota, \quad D = m'\Sigma^{-1}\iota, \quad E = m'\Sigma_s^{-1}m. \tag{4.14}$$

Various approaches exist to construct the signal m for the mean and to choose the target return. Here, I follow Engle et al. (2019) and Moura et al. (2020) and construct m using the momentum factor of Jegadeesh and Titman (1993). For each of the n stocks the individual momentum m_i, $i = 1, \ldots, n$, is computed as the geometric average of the previous 252 returns, but excluding the 21 most recent returns. Collecting all the momentum in a vector yields the signal m. The target return is computed as the arithmetic average of the momentums of those stocks that belong to the top-quintile stocks ranked according to momentum. Also for these mean-variance with momentum strategies, I combine the problem with the norm constraints defined in Equation (4.11), setting $s = 0.3$ and $s = 0.0$ to construct three similar portfolios as in the GMVP exercise. The solutions on the exposure-constrained allocations are obtained by making use of the CVX toolbox of Grant and Boyd (2014). For evaluating the performance, I use the empirical Sharpe ratio defined as

$$\hat{\theta} = \frac{\bar{R}_p}{\sqrt{\hat{\sigma}^2}},$$

with a superior forecasting performance indicated by larger values of the Sharpe ratio. The Sharpe ratio is considered among the most important economic portfolio evaluation criteria for not infinitely risk averse investors. The results provided in Table C.1 in Appendix C.4 can be summarized as follows: In the unrestricted allocation, the ATVGL again performs best

among all considered approaches as indicated by the largest Sharpe ratio and is also the best performing approach among all strategies based on the selection given in Equations (4.12) - (4.14) in the restricted allocations. However, for the gross-exposure constrained portfolios, the naïve allocation rule which does not depend on an estimation of the mean-return, leads to the largest Sharpe ratios which demonstrates once again the poor signal-to-noise ratio for daily return observations that make estimation of the mean return notoriously hard.

As a second robustness check, I conduct all analyses also for cross-sectional dimension $n = 200$ and provide the results in Tables C.2 – C.4 in Appendix C.4. Along the unrestricted GMVP predictions, ATVGL is again clearly the best performing approach: It is the only model to be included in the 90% MCS for the full out-of-sample window and is also the only one to be included in the model confidence sets in all five-year subsamples. Although the concentration ratio is two times as high as for the analysis with $n = 100$, the short window estimation SHR-nl* leads again to overall lower variances compared to SHR-nl. However, DCC-nl is in this setting the benchmark with the best performance on average as well as in six out of the eight subsamples. It can be assumed that DCC can cope exceptionally well with an extended asset universe as this leaves the estimation of the univariate volatility dynamics unaffected. The exposure-constrained GMVP results are qualitatively broadly similar compared to the setting with $n = 100$, but with DCC-nl also distinctly better than the remaining benchmarks. In the long-only exercise it even leads to significantly lower variance than ATVGL. Turning to the mean-variance with momentum signal-portfolio, I observe that in the restricted and unrestricted exercise, ATVGL leads to the highest Sharpe ratio among all models except for the naïve allocation, which, however, outperforms the proposed approach only in the 100/0-strategy. Summing up, the application to $n = 200$ confirms that ATVGL is well applicable also to settings with larger asset universes and is better able to cope with these than the static approaches. With exception of the very rigid long-only strategies, it is again the best performing model compared to a long list of very successful static and dynamic benchmarks, including simple selection strategies like the equally weighted allocation, static and dynamic shrinkage approaches as well as models incorporating factor structures (SHR-1f and AFM-DCC-nl).

4.5. Conclusion

In this paper, I propose to use a time-varying graphical LASSO approach using the alternating direction method of multipliers to model inverse covariance matrices in a large-dimensional system of financial assets. In particular, I consider daily returns on U.S.-traded stocks. Empirical results show that the optimal degree of model flexibility is not stable over time and that the estimates of unconditional covariance exhibit high sensitivity to the length of the estimation window. My proposed approach, Augmented TVGL with dynamic calibration, addresses these issues: it represents a new way to induce data-driven sparsity in financial portfolio optimization. The adaptive calibration scheme delivers uniformly good

performance over long time horizons and for different cross-sectional dimensions. By choosing particular penalty functions, the ATVGL is flexible to accommodate specific characteristics of the problem, such as minimum variance allocation predictions. Specific requirements, such as gross exposure constraints, can be easily incorporated. Notably, although clearly all empirical analyses depend on the stocks considered, due to the large underlying data base and the asset selection based on market capitalization, it can be assumed that the obtained results are quite general and of interest for financial portfolio managers.

For future work, the approach could be extended in the context of portfolio optimization by introducing alternative penalties that, for example, provide explicit turnover control, or it could be applied to other sampling frequencies and other asset classes for which signals for mean returns were easier to derive so that they could also be accommodated in the penalty functions. Potentially, other financial problems could be addressed, e.g., modeling certain risk metrics such as value-at-risk or expected shortfall, for which the tail behavior of the joint distribution is very important. It can be assumed that for this purpose the abandonment of the Gaussian assumption for the joint vector of returns would be useful.

Appendix A.

Appendix for Chapter 2

A.1. Elicitability of GMVP weights

We begin with stating an auxiliary result. While the result is straightforward, we are not aware of a reference and thus provide a proof.

Lemma 1. *A mixture of two $n-$variate distributions with mean vectors μ_a, μ_b and covariance matrices Σ_a, Σ_b has precision matrix*

$$\Sigma_\pi^{-1} = \frac{\pi(1-\pi)}{1 + \pi(1-\pi)d'V_\pi^{-1}d} \; dd'V_\pi^{-1},$$

where $\pi \in [0,1]$ is the mixture probability for the first component, and

$$
\begin{aligned}
d &= (\mu_a - \mu_b), \\
V_\pi &= \pi\Sigma_a + (1-\pi)\Sigma_b.
\end{aligned}
$$

Proof. The covariance matrix of the mixture is given by

$$\Sigma_\pi = V_\pi + (\mu_a, \mu_b) \begin{pmatrix} \pi(1-\pi) & -\pi(1-\pi) \\ -\pi(1-\pi) & \pi(1-\pi) \end{pmatrix} \begin{pmatrix} \mu'_a \\ \mu'_b \end{pmatrix}.$$

The result then follows from a variant of the Woodbury matrix identity (Petersen and Pedersen, 2012, Equation 159). $\qquad\square$

To discuss elicitability, we next introduce some notation. Let $\mathcal{G}$ denote the family of n-variate continuous distributions with finite mean vector and covariance matrix. For a typical member G of this family, we denote the covariance matrix associated with G by $\Sigma(G)$. The $(n-1) \times 1$ vector $\beta_{1:n-1}(G)$ contains the associated GMVP weights for the first $n-1$ assets (the remaining weight is implied by the constraint that the weights sum to unity). Moreover, the $n \times 1$ vector $\beta(G)$ is given by $(\beta_0(G), \beta_{1:n-1}(G)')'$, where the first element is the expected GMVP return implied by G.

Proposition 2. $\beta_{1:n-1}$ *does not have convex level sets. That is, a convex combination G_π*

of two distributions $G_a, G_b \in \mathcal{G}$ such that $\beta_{1:n-1}(G_a) = \beta_{1:n-1}(G_b) = b_{1:n-1}$ generally has $\beta_{1:n-1}(G_\pi) \neq b_{1:n-1}$.

Proof. Follows from Lemma 1. As a simple example, the violation can be checked for the case $n = 2$, $\Sigma(G_a) = \Sigma(G_b) = I_2$, $\mu(G_a) = (1, 1)'$ and $\mu(G_b) = (2, 2)'$. □

Since $\beta_{1:n-1}$ does not have convex level sets, it can not be elicitable. The necessity of convex level sets for elicitability is formally stated for the case of a univariate predictand in Gneiting (2011, Theorem 6). However, the condition is also necessary in the present case of an $n - 1$ variate predictand, c.f. the proof of Lemma 1 in Lambert et al. (2008) as well as the discussion by Fissler and Ziegel (2019, p. 1170).

Since β is elicitable, it must have convex level sets. For example, consider two distributions $G_a, G_b \in \mathcal{G}$ such that $\Sigma(G_a) \propto \Sigma(G_b)$ and $\beta_0(G_a) = \beta_0(G_b)$. This setup implies that $\beta(G_a) = \beta(G_b)$. Using Lemma 1 and the fact that

$$\Big(\mu(G_a) - \mu(G_b)\Big)' \Sigma(G_a)^{-1} \iota_n = 0,$$

it can be shown that $\Sigma(G_\pi)^{-1} \iota_n \propto \Sigma(G_a)^{-1} \iota_n$ and hence $\beta(G_\pi) = \beta(G_a) = \beta(G_b)$, which illustrates that β has convex level sets.

A.2. Invariance of the GMVP models

A.2.1. Preliminaries

In order discuss the invariance of the dynamic GMVP models based on the RLS recursion (in Section 2.3.1) and the GAS specifications (in Section 2.3.2), we begin with stating the relationship between the auxiliary regressions identifying the GMVP weights for different baseline assets.

When asset n is used as baseline asset, then the GMVP auxiliary regression (as reproduced from Section 2.2.1) is

$$Y_t = X_t'\beta_t + \varepsilon_t, \tag{A.1}$$

$$Y_t = R_{nt}, \qquad X_t' = (1, R_{nt} - R_{1t}, \dots, R_{nt} - R_{n-1t}),$$

with expected GMVP return β_{0t} and GMVP weights

$$\omega_{it}^* = \begin{cases} \beta_{it} & i = 1, \dots, n-1, \\ 1 - \iota_{n-1}'\beta_{1:n-1t} & i = n. \end{cases}$$

Suppose we select a different baseline asset, say asset k instead of asset n. Then the

corresponding variables in the GMVP regression of the form (A.1) are given by

$$\tilde{Y}_t = R_{kt}, \qquad \tilde{X}'_t = (1, R_{kt} - R_{1t}, \ldots, R_{kt} - R_{k-1t}, R_{kt} - R_{nt},$$
$$R_{kt} - R_{k+1t}, \ldots, R_{kt} - R_{n-1t}).$$

These regression variables associated with baseline asset k obtain from those for baseline asset n according to the one-to-one transformation

$$\tilde{Y}_t = d'_k X_t + Y_t, \qquad \tilde{X}_t = R_k X_t, \tag{A.2}$$

where

$$d'_k = (0, -e'_k), \qquad R_k = \begin{pmatrix} 1 & 0'_{n-1} \\ 0_{n-1} & S_k \end{pmatrix},$$

$$S_k = \begin{pmatrix} I_{k-1} & -\iota_{k-1} & 0_{k-1 \times n-k-1} \\ 0'_{k-1} & -1 & 0'_{n-k-1} \\ 0_{n-k-1 \times k-1} & -\iota_{n-k-1} & I_{n-k-1} \end{pmatrix}.$$

Here we have used 0_ℓ to denote the ℓ-dimensional Null vector, $0_{\ell_1 \times \ell_2}$ to denote the $(\ell_1 \times \ell_2)$-dimensional Null matrix and e_k to denote the k'th column of the $(n-1)$-dimensional identity matrix I_{n-1}. For the permutation matrix R_k it holds that

$$R_k = R_k^{-1}, \qquad R'_k d_k = -d_k. \tag{A.3}$$

Using the equalities $Y_t = \tilde{Y}_t + d'_k \tilde{X}_t$ and $X_t = R_k \tilde{X}_t$ resulting from Equations (A.2) and (A.3) in the GMVP regression (A.1) for baseline asset n, yields the following equivalent GMVP regression for baseline asset k:

$$\tilde{Y}_t = \tilde{X}'_t \tilde{\beta}_t + \varepsilon_t, \tag{A.4}$$

with

$$\tilde{\beta}_t = R'_k \beta_t - d_k, \tag{A.5}$$

$$\omega^*_{it} = \begin{cases} \tilde{\beta}_{it} & i = 1, \ldots, k-1, k+1, \ldots, n-1 \\ \tilde{\beta}_{kt} & i = n \\ 1 - \iota'_{n-1} \tilde{\beta}_{1:n-1t} & i = k, \end{cases}$$

and expected portfolio return $\tilde{\beta}_{0t}$. An immediate implication of Equation (A.5) is that the coefficients β_t in the GMVP regression used to identify the GMVP weights under baseline asset n are in one-to-one correspondence to the coefficients $\tilde{\beta}_t$ in the GMVP regression associated with baseline asset k equivalently identifying the GMVP weights. Moreover,

Equation (A.4) shows that the error terms of the GMVP regressions for both baseline assets are the same. Whence the value of the loss function (2.4) computed for (Y_t, X_t, β_t) is the same as that computed for $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$.

With the one-to-one mapping between (Y_t, X_t, β_t) and $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$ as given by Equations (A.2) and (A.5) we can now discuss the invariance of our proposed dynamic GMVP models w.r.t. the selection of the baseline asset. The following conditions are necessary and sufficient for this invariance: First, there must be a parametrization (including initial conditions) for the model associated with baseline asset k (for any $k \neq n$) that leads to the same predictions for the GMVP weights as the model associated with baseline asset n which requires according to Equation (A.5) that the predictions satisfy $\tilde{\beta}_{t+1} = R_k' \beta_{t+1} - d_k \ \forall t$; second, this parametrization for baseline asset k must be in one-to-one correspondence with that for baseline asset n. Since for $\tilde{\beta}_t = R_k' \beta_t - d_k$ the loss function for (Y_t, X_t, β_t) is the same as for $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$ it follows that under those two conditions the M-estimator as defined in Equation (2.8) is invariant w.r.t. the choice of the baseline asset and leads to the same estimates for the predicted GMVP weights. The one-to-one correspondence of the parameterizations means, for example, that if the matrices A and B of the GAS recursion (2.21) have a diagonal form for baseline asset n, then the reparametrization of those two matrices, which gives the same predictions for the weights, must also be diagonal. The same applies for the initial condition $\mathbb{E}_0[\tilde{X}_1 \tilde{X}_1']$ (see Equation 2.23).

A.2.2. Invariance of the RLS models

In Lemma 2 we provide the parametrization for the RLS-EF model of Section 2.3.1 for baseline asset k which leads to the same GMVP predictions as the RLS-EF model for baseline asset n.

Lemma 2. *Consider the RLS-EF model for baseline asset n given by*

$$\beta_{t+1} \ = \ \beta_t + \Omega_t^{-1} X_t (Y_t - X_t' \beta_t), \tag{A.6}$$

$$\Omega_t \ = \ X_t X_t' + \lambda \Omega_{t-1}, \tag{A.7}$$

with initial conditions β_1 and Ω_0. This RLS-EF model is equivalent to the following RLS-EF model for baseline asset k:

$$\tilde{\beta}_{t+1} \ = \ \tilde{\beta}_t + \tilde{\Omega}_t^{-1} \tilde{X}_t (\tilde{Y}_t - \tilde{X}_t' \tilde{\beta}_t), \tag{A.8}$$

$$\tilde{\Omega}_t \ = \ \tilde{X}_t \tilde{X}_t' + \tilde{\lambda} \tilde{\Omega}_{t-1}, \tag{A.9}$$

with

$$\tilde{\lambda} = \lambda, \tag{A.10}$$

$$\tilde{\beta}_1 = R'_k \beta_1 - d_k, \tag{A.11}$$

$$\tilde{\Omega}_0 = R_k \Omega_0 R'_k. \tag{A.12}$$

Specifically, for the β_t's obtained according to Equations (A.6) and (A.7) and the $\tilde{\beta}_t$'s obtained according to Equations (A.8)-(A.12) it holds that $\tilde{\beta}_t = R'_k \beta_t - d_k$ for all $t = 2, 3, \ldots$, as desired.

Proof. From (A.9) with (A.10) and (A.12) it follows that

$$R_k \tilde{\Omega}_t R'_k = \Omega_t, \quad t = 0, 1, 2, \ldots. \tag{A.13}$$

Using (A.2), (A.10) and (A.13) together with $\tilde{\beta}_t = R'_k \beta_t - d_k$ on the r.h.s. of Equation (A.8) shows that if $\tilde{\beta}_t = R'_k \beta_t - d_k$ and β_{t+1} and $\tilde{\beta}_{t+1}$ are generated according to Equations (A.6) and (A.8), then it holds that $\tilde{\beta}_{t+1} = R'_k \beta_{t+1} - d_k$. This combined with the initial condition $\tilde{\beta}_1 = R'_k \beta_1 - d_k$ in (A.11) completes the proof. $\square$

Equation (A.10) shows that the parameter λ associated with baseline asset n is in (trivial) one-to-one correspondence with the parameter $\tilde{\lambda}$ associated with baseline asset k. So if the initial conditions $(\tilde{\beta}_1, \tilde{\Omega}_0)$ for baseline asset k, as defined by (A.11) and (A.12), are also in a one-to-one correspondence with the initial conditions (β_1, Ω_0) selected for the baseline asset n, then Lemma 2 implies that the RLS-EF model is invariant w.r.t. the choice of the baseline asset.

Our selection for β_1 and Ω_0 are (as reproduced from Section 2.3.1)

$$\beta_1 = (0, \iota'_{n-1}/n)', \quad \Omega_0 = \gamma \begin{pmatrix} 1 & 0'_{n-1} \\ 0_{n-1} & \hat{\sigma}_R^2 (1 - \hat{\rho}_R) C \end{pmatrix}, \tag{A.14}$$

where $C = I_{n-1} + \iota_{n-1} \iota'_{n-1}$ and $\hat{\sigma}_R^2$ and $\hat{\rho}_R$ are scalars that are independent from the selected baseline asset. The transformation of those initial conditions according to the r.h.s. of (A.11) and (A.12) yields

$$R'_k \beta_1 - d_k = \beta_1, \quad R_k \Omega_0 R'_k = \Omega_0,$$

for any γ, so that $\tilde{\beta}_1 = \beta_1$ and $\tilde{\Omega}_0 = \Omega_0$ with $\tilde{\gamma} = \gamma$. This implies that there is a one-to-one correspondence between the initial conditions for baseline asset k and those for baseline asset n, as required for the invariance of the RLS-EF model.

This proof of invariance of the RLS-EF model in Section (2.3.1) can straightforwardly be adapted to show the invariance of the RLS-REF model in Section (2.3.1).

A.2.3. Invariance of the GAS model

Lemma 3 provides the parametrization for the GAS model of Section 2.3.2 for baseline asset k which leads to the same GMVP predictions as the GAS model for baseline asset n.

Lemma 3. *Consider the GAS model in Equations (2.21) and (2.23) for baseline asset n written as*

$$\beta_{t+1} = c + B\beta_t - A(\mathbb{E}_{t-1}[X_t X_t'])^{-1} X_t (Y_t - X_t'\beta_t), \tag{A.15}$$

$$\mathbb{E}_t[X_{t+1} X_{t+1}'] = \kappa \mathbb{E}_{t-1}[X_t X_t'] + (1-\kappa) X_t X_t', \tag{A.16}$$

with initial condition β_1 and $\mathbb{E}_0[X_1 X_1']$. This GAS model is equivalent to the following GAS model for baseline asset k:

$$\tilde{\beta}_{t+1} = \tilde{c} + \tilde{B}\tilde{\beta}_t - \tilde{A}(\mathbb{E}_{t-1}[\tilde{X}_t \tilde{X}_t'])^{-1} \tilde{X}_t (\tilde{Y}_t - \tilde{X}_t'\tilde{\beta}_t), \tag{A.17}$$

$$\mathbb{E}_t[\tilde{X}_{t+1} \tilde{X}_{t+1}'] = \tilde{\kappa} \mathbb{E}_{t-1}[\tilde{X}_t \tilde{X}_t'] + (1-\tilde{\kappa}) \tilde{X}_t \tilde{X}_t', \tag{A.18}$$

with

$$\tilde{\kappa} = \kappa \tag{A.19}$$

$$\tilde{c} = R_k'(c + d_k - Bd_k), \tag{A.20}$$

$$\tilde{B} = R_k' B R_k', \tag{A.21}$$

$$\tilde{A} = R_k' A R_k', \tag{A.22}$$

$$\tilde{\beta}_1 = R_k'\beta_1 - d_k, \tag{A.23}$$

$$\mathbb{E}_0[\tilde{X}_1 \tilde{X}_1'] = R_k \mathbb{E}_0[X_1 X_1'] R_k'. \tag{A.24}$$

Specifically, for the β_t's obtained according to Equations (A.15)-(A.16) and the $\tilde{\beta}_t$'s obtained according to Equations (A.17)-(A.24) it holds that $\tilde{\beta}_t = R_k'\beta_t - d_k$ for all $t = 2, 3, \ldots$, as desired.

Proof. From Equations (A.18) with (A.19) and (A.24) it follows that

$$R_k \mathbb{E}_t[\tilde{X}_{t+1} \tilde{X}_{t+1}'] R_k' = \mathbb{E}_t[X_{t+1} X_{t+1}'], \quad t = 0, 1, 2, \ldots. \tag{A.25}$$

Using (A.2), (A.20)-(A.22) and (A.25) together with $\tilde{\beta}_t = R_k'\beta_t - d_k$ on the r.h.s. of Equation (A.17) shows that if $\tilde{\beta}_t = R_k'\beta_t - d_k$ and β_{t+1} and $\tilde{\beta}_{t+1}$ are generated according to Equations (A.15) and (A.17), then it holds that $\tilde{\beta}_{t+1} = R_k'\beta_{t+1} - d_k$. This combined with the initial condition $\tilde{\beta}_1 = R_k'\beta_1 - d_k$ in (A.23) completes the proof. $\square$

Our proposed GAS model imposes the parameter restrictions (reproduced from Section

2.3.2)

$$c = (I_n - B) \begin{pmatrix} 0 \\ \iota_{n-1}/n \end{pmatrix}, \tag{A.26}$$

$$B = \begin{pmatrix} b_0 & 0'_{n-1} \\ 0_{n-1} & b_1 I_{n-1} \end{pmatrix}, \quad A = \begin{pmatrix} a_0 & 0'_{n-1} \\ 0_{n-1} & a_1 I_{n-1} \end{pmatrix}, \tag{A.27}$$

and uses the initial conditions

$$\beta_1 = (0, \iota'_{n-1}/n)', \quad \mathbb{E}_0[X_1 X'_1] = \begin{pmatrix} 1 & 0'_{n-1} \\ 0_{n-1} & \hat{\sigma}_R^2(1 - \hat{\rho}_R)C \end{pmatrix}, \tag{A.28}$$

where $C = I_{n-1} + \iota_{n-1}\iota'_{n-1}$ and $\hat{\sigma}_R^2$ and $\hat{\rho}_R$ are scalars that are independent from the selected baseline asset. According to Lemma 3, its invariance w.r.t. the selection of the baseline asset requires that $\tilde{c}$, $\tilde{B}$, $\tilde{A}$, $\tilde{\beta}_1$, and $\mathbb{E}_0[\tilde{X}_1 \tilde{X}'_1]$, which result from the transformations (A.20)-(A.24) from c, B, A, β_1, and $\mathbb{E}_0[X_1 X'_1]$ given in (A.26)-(A.28), are each in a one-to-one correspondence.

As it is easy to show, those transformations yield

$$R'_k(c + d_k - Bd_k) = c, \quad R'_k B R'_k = B, \quad R'_k A R'_k = A,$$

for any (b_0, b_1, a_0, a_1), and

$$R'_k \beta_1 - d_k = \beta_1, \quad R_k \mathbb{E}_0[X_1 X'_1] R'_k = \mathbb{E}_0[X_1 X'_1].$$

So this together with $(A.19)$ implies that all parameters and initial conditions $(\kappa, c, B, A, \beta_1, \mathbb{E}_0[X_1 X'_1])$ associated with baseline asset n are in one-to-one correspondence to the parameters and initial conditions $(\tilde{\kappa}, \tilde{c}, \tilde{B}, \tilde{A}, \tilde{\beta}_1, \mathbb{E}_0[\tilde{X}_1 \tilde{X}'_1])$ associated with baseline asset k, as required for the invariance of the GAS model.

A.3. Data and additional results

A.3.1. Data

Figure A.1 plots the times series of the daily returns on the equally weighted portfolio consisting of the 1000 largest stocks of our data set, which have a complete series of reported returns over the full sample period (01/02/2002 to 12/09/2019). The size of the stocks are measured in terms of market capitalization. The grey shaded area in Figure A.1 marks the out-of-sample window used for our out-of-sample forecasting experiments in Section 2.4, which ranges from 01/03/2007 to 12/09/2019.
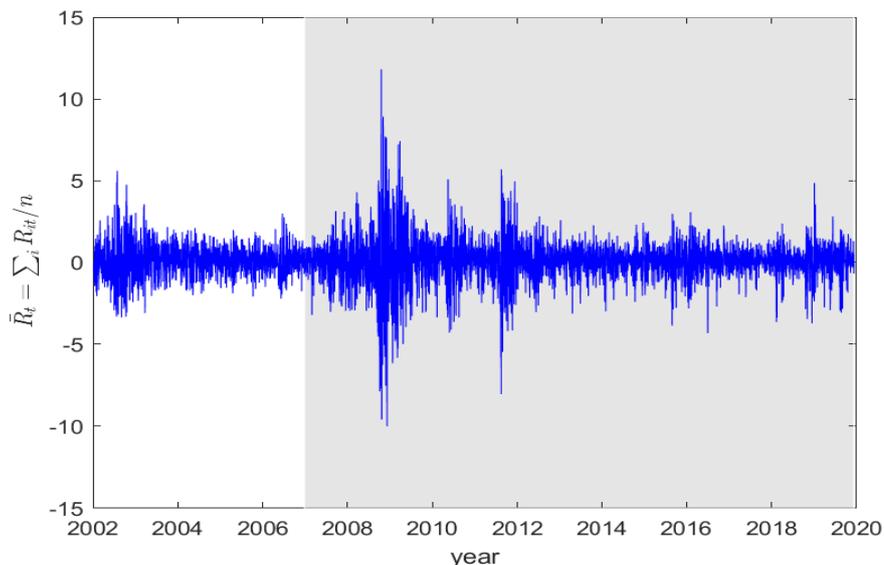
Figure A.1.: Time plot of the daily returns on the equally weighted portfolio including the 1000 largest stocks.

A.3.2. RLS information matrix Ω_t in the classical and regularized exponential forgetting

Here we provide the bounds of the information matrix Ω_t of the RLS-EF with classical exponential forgetting and the RLS-REF with regularized exponential forgetting. (For further details, see Kulhavỳ and Zarrop, 1993, Shin and Lee, 2020).

For this we use the notation $V \geq W$ $(V > W)$ to indicate that $V - W$ is a positive semi-definite (positive definite) matrix and 0 and ∞ stand for matrices with zero and infinite eigenvalues. Assuming that the return differences in X_t observed in a given empirical sample are bounded such that $0 \leq X_t X_t' \leq C < \infty$, $t = 1, 2, \ldots$, then Ω_t in the RLS-EF, given by $\Omega_t = X_t X_t' + \lambda \Omega_{t-1}$, is bounded from below and above as

$$0 \leq \Omega_t \leq \Omega_0 + \frac{1}{1 - \lambda} C, \quad t = 1, 2, \ldots . \tag{A.29}$$

The regularized Ω_t in the RLS-REF, given by $\Omega_t = X_t X_t' + \lambda \Omega_{t-1} + (1 - \lambda)\Omega_0$, is bounded from below and above as

$$\Omega_0 \leq \Omega_t \leq \Omega_0 + \frac{1}{1 - \lambda} C, \quad t = 1, 2, \ldots . \tag{A.30}$$

In order to illustrate the effect of the regularization on the time series behavior of the RLS information matrix Ω_t, we provide in Figure A.2 time series plots of its logarithmic eigenvalues that result from fitted RLS-EF and RLS-REF models for portfolio sizes $n \in \{50, 200, 500, 1000\}$. The data used are the same as those for the empirical application
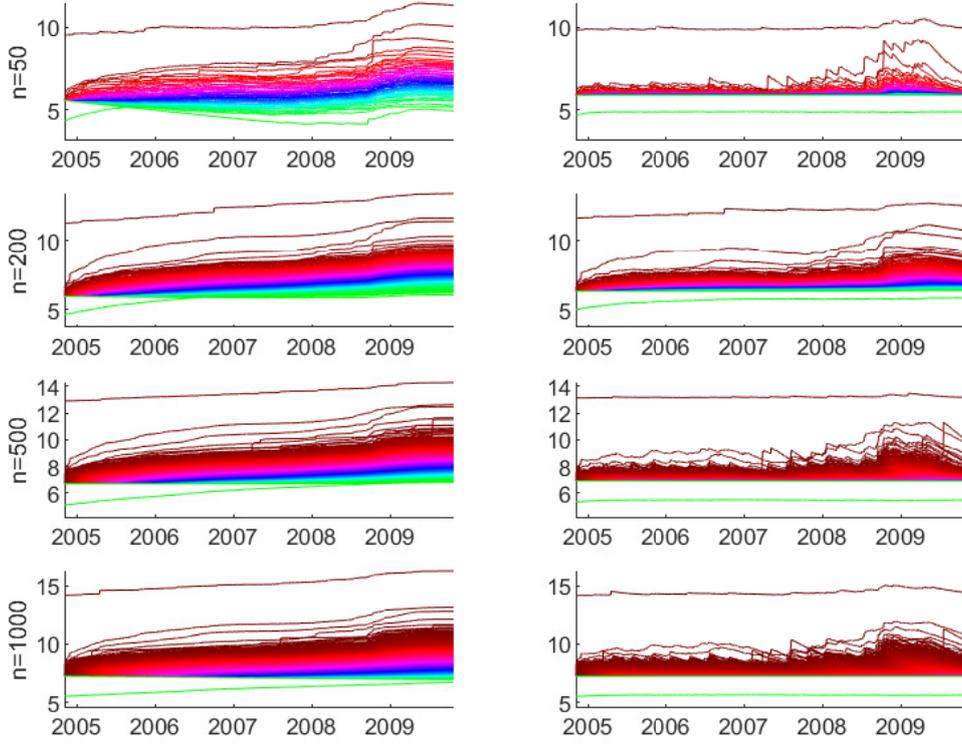
Figure A.2.: Time plot of the logarithmic eigenvalues of Ω_t that result from fitted RLS-EF models (left panels) and RLS-REF models (right panels) for portfolio sizes $n \in \{50, 200, 500, 1000\}$. The NLS parameter estimates in the RLS-EF for the different n's are $(n, \hat{\lambda}, \hat{\gamma}) = \{(50, 0.9977, 77), (200, 0.9998, 105), (500, 1.0000, 166), (1000, 0.9998, 249)\}$ and for the RLS-REF $(n, \hat{\lambda}, \hat{\gamma}) = \{(50, 0.9692, 104), (200, 0.9954, 151), (500, 0.9802, 206), (1000, 0.9840, 246)\}$.

discussed in Section 2.4 and the estimation window ranges from 11/11/2004 to 10/23/2009 covering $T = 1250$ trading days.

The time plots show that the regularization of Ω_t in the RLS-REF stabilizes the behavior of its largest eigenvalues, although (for a given λ) the upper bound of Ω_t in the RLS-REF is the same as for the RLS-EF (see Equations A.29 and A.30). This stabilizing effect in the most strongly excited directions is essentially achieved by the fact that the estimates for λ in the RLS-REF are smaller than in the RLS-EF, while in the poorly exited directions the eigenvalues of the RLS-REF are bounded by the prior lower bound. The parameter estimates are provided in the caption of Figure A.2.

A.3.3. Superior predictive ability test results

As a robustness check for the MCS results presented in Section 2.4, Table A.1 reports p-values of the test for superior predictive ability (SPA). The test's hypothesis is that a given

Table A.1.: Hansen (2005) test for superior predictive ability results

	$T = 250$				$T = 1250$			
n	50	200	500	1000	50	200	500	1000
n/T	0.20	0.80	2.00	4.00	0.04	0.16	0.40	0.80
RLS-EF	0.40	0.65	0.00	0.00	0.04	0.01	0.47	0.06
RLS-REF	0.99	0.99	0.99	0.29	1.00	1.00	0.64	0.13
GAS	0.02	0.01	0.00	0.00	0.03	0.06	0.02	0.00
DCC-s	0.00	0.00	–	–	0.00	0.00	0.00	0.00
DCC-nl	0.01	0.06	0.58	1.00	0.00	0.02	1.00	1.00
OLS	0.00	0.00	–	–	0.00	0.00	0.00	0.00
SHR-l	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SHR-nl	0.06	0.13	0.32	0.00	0.00	0.00	0.01	0.00
naïve	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: p-values of the Hansen (2005) test for superior predictive ability. Rows correspond to benchmark models, columns correspond to forecast experiments. The one-sided hypothesis is that the benchmark is at least as good as all competitors.

benchmark model is at least as good as all of its competitors. Hence a small p-value yields evidence against the benchmark model. Our implementation is based on the stationary bootstrap with a mean block length of $\lfloor T_{eval}^{1/3} \rfloor = \lfloor 3257^{1/3} \rfloor = 14$, and we use the conservative estimator $\hat{\mu}^u$ for the mean loss differential (Hansen, 2005, p. 372).

The results in Table A.1 are qualitatively very similar to the results based on the MCS approach. In particular, successful models that are included in the MCS typically achieve high SPA p-values when used as a benchmark model. Conversely, for poor performing models that are not included in the MCS, the SPA p-value is typically very small. For example, the RLS-REF model, which is contained in the MCS for all eight experiments, generally achieves an SPA p-value exceeding ten percent. In this regard, RLS-REF outperforms all other models, each of which in at least one of the empirical experiments has a SPA p-value that is less than one percent. By contrast, the naïve model, which is not part of the MCS for any of the eight forecast experiments, attains an SPA p-value of zero in each case.

A.3.4. Sharpe ratios

In this section we provide the out-of-sample Sharpe ratios of the predicted GMVP allocations for our proposed GMVP models and the benchmark models used in Section 2.4. The Sharpe ratio is defined as $\hat{\mathbb{E}}[R_{pt}]/\hat{\mathbb{V}}[R_{pt}]^{1/2}$, where $\hat{\mathbb{E}}[R_{pt}]$ is the time average of the portfolio returns for the predicted GMVP allocations and $\hat{\mathbb{V}}[R_{pt}]$ the corresponding variance.

The results for the eight (n/T)-scenarios considered in Section 2.4 are reported in Table A.2. They show that the ranking of the models for the various (n/T)-cases is quite different. However, the GAS model and the RLS-EF each perform best in two cases and the RLS-REF in one case.

Table A.2.: Out-of-sample GMVP Sharpe ratio (×10)

	$T = 250$				$T = 1250$			
n	50	200	500	1000	50	200	500	1000
n/T	0.20	0.80	2.00	4.00	0.04	0.16	0.40	0.80
RLS-EF	**0.704**	0.644	0.573	0.661	0.590	**0.616**	0.646	0.601
RLS-REF	0.688	0.627	**0.711**	0.651	0.549	0.544	0.614	0.627
GAS	0.424	0.308	0.512	**0.919**	0.428	0.485	**0.840**	0.599
DCC-s	0.629	0.466	–	–	**0.591**	0.560	0.543	0.449
DCC-nl	0.650	**0.650**	0.636	0.544	0.588	0.602	0.670	0.680
OLS	0.655	0.418	–	–	0.494	0.550	0.477	0.328
SHR-l	0.662	0.488	0.580	0.659	0.495	0.555	0.485	0.386
SHR-nl	0.659	0.600	0.615	0.621	0.492	0.596	0.640	**0.682**
naïve	0.345	0.352	0.338	0.328	0.345	0.357	0.345	0.328

Note: Sharpe ratio (multiplied by 10) of the predicted GMVP portfolios for portfolio sizes $n \in \{50, 200, 500, 1000\}$. Parameter estimation is based on a sample of length T. Bold numbers indicate largest Sharpe ratio across all models.

A.3.5. Parameter estimates

In Figures (A.3)-(A.5) we provide histograms of the NLS estimates for the parameters of the proposed dynamic GMVP models, the RLS-EF (Figure A.3), the RLS-REF (Figure A.4), and the GAS model (Figure A.5). These are the parameter estimates which are obtained in our out-of-sample forecasting experiment in Section 2.4.2, where the parameters are re-estimated every month on a rolling estimation window scheme with window lengths, $T = 250$ and $T = 1250$ days. The estimates presented in Figures (A.3)-(A.5) refer to the portfolio sizes $n = 200$ and $n = 1000$.
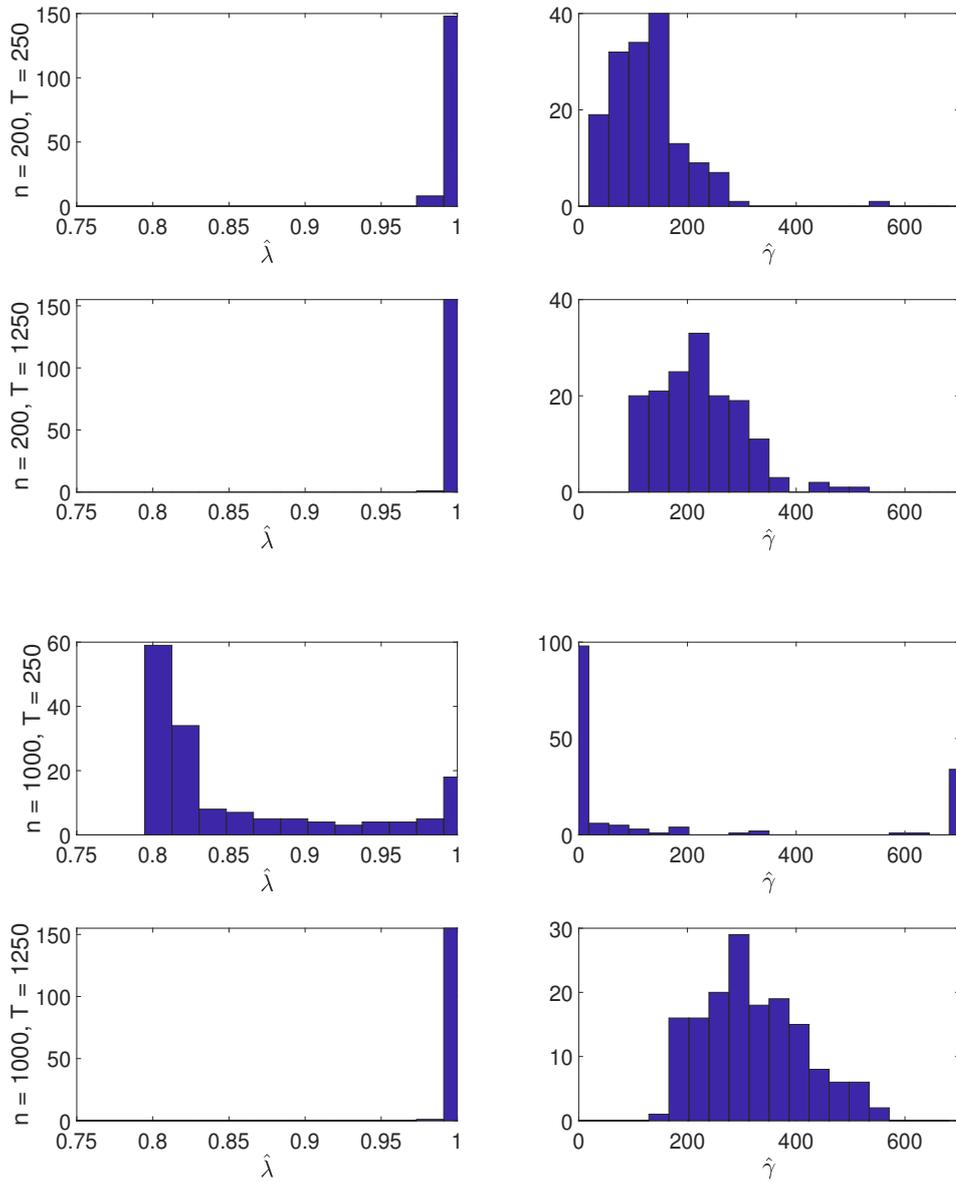
Figure A.3.: Histogram of the NLS-estimates for the RLS-EF model parameters λ (left panels) and γ (right panels) for portfolio sizes $n = 200$ (upper four panels) and $n = 1000$ (lower four panels); Estimation sample sizes are $T = 250$ and $T = 1250$.
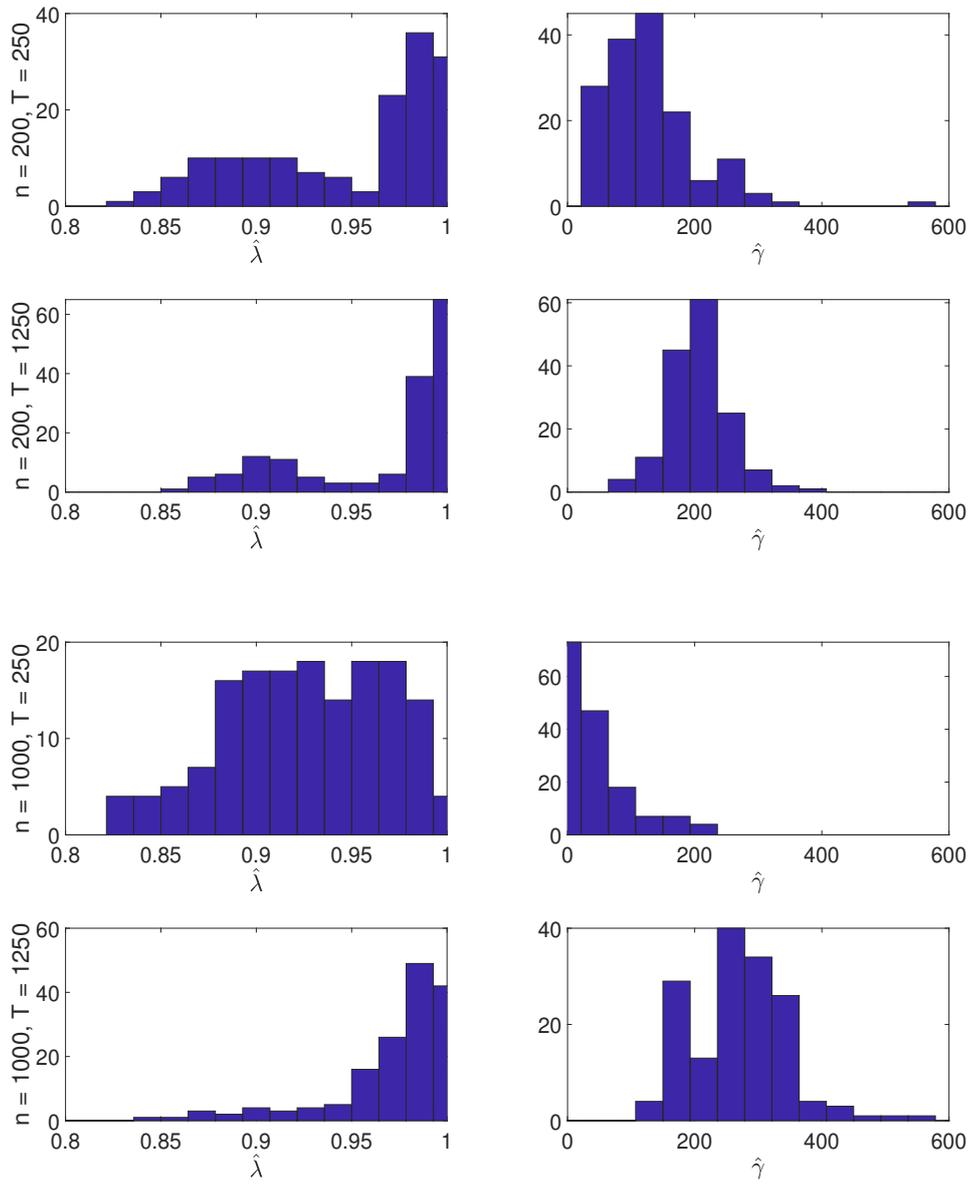
Figure A.4.: Histogram of the NLS-estimates for the RLS-REF model parameters λ (left panels) and γ (right panels) for portfolio sizes $n = 200$ (upper four panels) and $n = 1000$ (lower four panels); Estimation sample sizes are $T = 250$ and $T = 1250$.
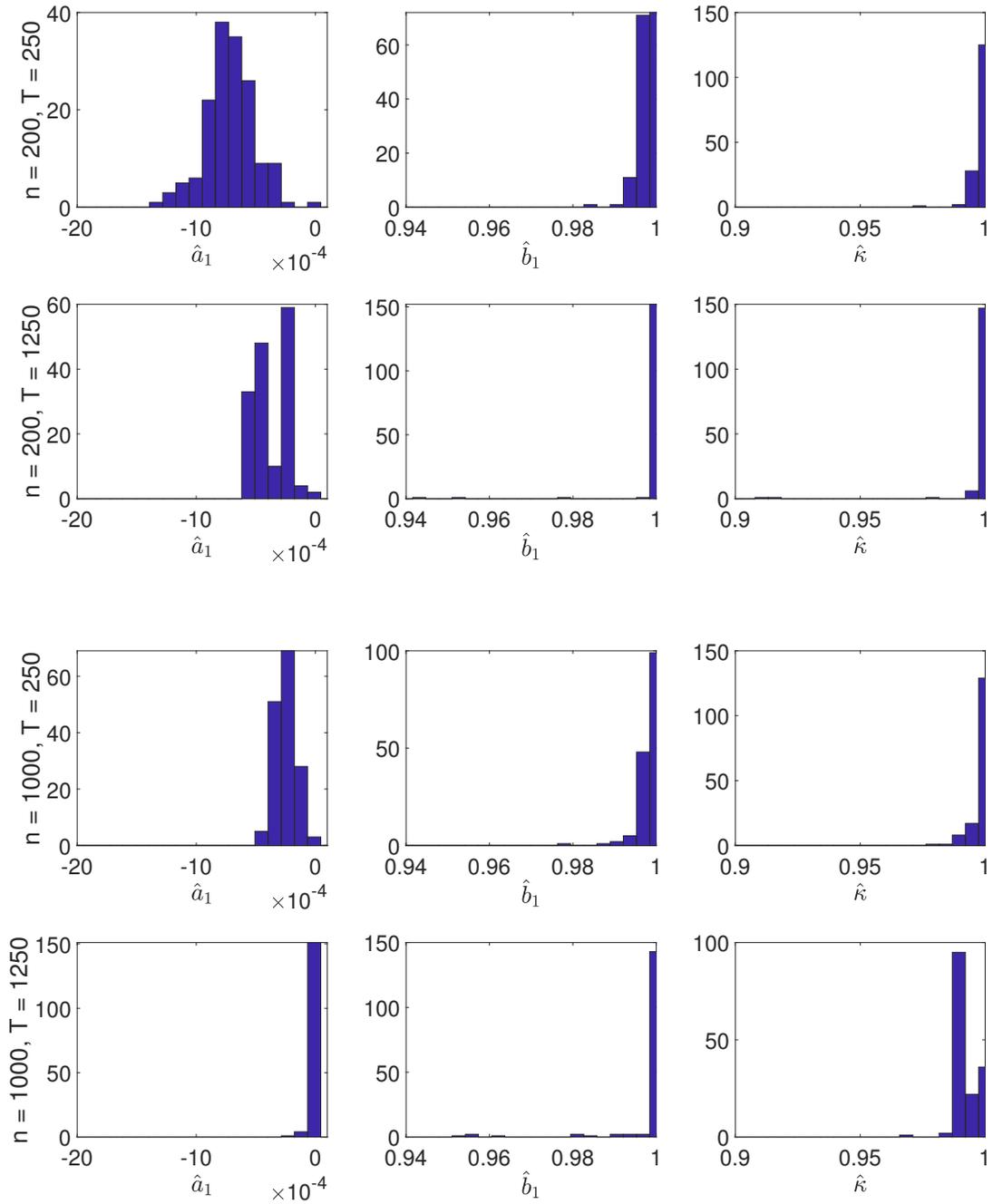
Figure A.5.: Histogram of the NLS-estimates for the GAS model parameters a_1 (left panels), b_1 (middle panels) and κ (right panels) for portfolio sizes $n = 200$ (upper four panels) and $n = 1000$ (lower four panels); Estimation sample sizes are $T = 250$ and $T = 1250$.

A.4. Statistical inference based on the M-estimator

For estimating the parameters θ of our proposed dynamic GMVP models specifying a function for $\beta_t = \beta(\theta, Z_{t-1}) = \beta_t(\theta)$ we use the NLS estimator as given in Equation (2.8), repeated here for convenience:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} L\left(\beta_t(\theta), R_t\right), \quad L\left(\beta_t(\theta), R_t\right) = [Y_t - X_t'\beta_t(\theta)]^2. \qquad (A.31)$$

The nonlinear function $\beta_t(\theta)$ depends on the corresponding model specification. For solving the minimization problem in Equation (A.31) we use the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm.

Under standard regularity conditions for M-estimators such as those of Amemiya (1985, Theorems 4.1.1 and 4.1.3), the NLS estimator $\hat{\theta}$ in Equation (A.31) is consistent and asymptotically normal with

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, D^{-1}WD^{-1}), \qquad (A.32)$$

where

$$D = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right], \quad W = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[g_t(\theta)g_t(\theta)'\right], \qquad (A.33)$$

and $g_t(\theta) = \partial L(\beta_t(\theta), R_t)/\partial\theta$ is the gradient of the loss function[1].

The asymptotic covariance matrix of the M-estimator as given by Equations (A.32)-(A.33) comprises the expected outer-product of the gradient of the loss function $\mathbb{E}[g_t(\theta)g_t(\theta)']$ and the expected Hessian $\mathbb{E}[\partial g_t(\theta)/\partial\theta']$.

The gradient of the GMVP loss function is given by

$$g_t(\theta) = \frac{\partial L(\beta_t(\theta), R_t)}{\partial \theta} = \frac{\partial \beta_t(\theta)'}{\partial \theta} \frac{\partial L(\beta_t(\theta), R_t)}{\partial \beta_t},$$

with

$$\frac{\partial L(\beta_t(\theta), R_t)}{\partial \beta_t} = \nabla_t = -2X_t[Y_t - X_t'\beta_t(\theta)],$$

so that its expected outer product obtains as

$$\mathbb{E}\left[g_t(\theta)g_t(\theta)'\right] = 4\mathbb{E}\left(\left[Y_t - X_t'\beta_t(\theta)\right]^2 \frac{\partial \beta_t(\theta)'}{\partial \theta} X_t X_t' \frac{\partial \beta_t(\theta)}{\partial \theta'}\right).$$

---

[1]Patton et al. (2019) provide specific sufficient conditions for the consistency and asymptotic normality (of the form as in Equation A.32) for an M-estimator similar to ours. As in their work, our estimator minimizes a consistent loss function for dynamically evolving parameters (β_t), which are given as functions that are measurable in $\mathcal{F}_{t-1}$ and continuous in the static parameters (θ).

Appendix A. Appendix for Chapter 2

The Hessian of the GMVP loss function is given by

$$\frac{\partial g_t(\theta)}{\partial \theta'} = \frac{\partial \beta_t(\theta)'}{\partial \theta} \frac{\partial \nabla_t}{\partial \theta'} + (\nabla_t' \otimes I_m) \frac{\partial \text{vec} \left[\partial \beta_t(\theta)'/\partial \theta \right]}{\partial \theta'},$$

where $\text{vec}(\cdot)$ denotes the operator which stacks the columns of a matrix into a vector, and $\otimes$ is the Kronecker matrix product. It follows that

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right] &= \mathbb{E}\left(\mathbb{E}_{t-1}\left[\frac{\partial \beta_t(\theta)'}{\partial \theta}\frac{\partial \nabla_t}{\partial \theta'} + (\nabla_t' \otimes I_m)\frac{\partial \text{vec}\left[\partial \beta_t(\theta)'/\partial \theta\right]}{\partial \theta'}\right]\right) \\
&= \mathbb{E}\left(\frac{\partial \beta_t(\theta)'}{\partial \theta}\mathbb{E}_{t-1}\left[\frac{\partial \nabla_t}{\partial \theta'}\right]\right) \\
&= 2\mathbb{E}\left(\frac{\partial \beta_t(\theta)'}{\partial \theta}X_t X_t'\frac{\partial \beta_t(\theta)}{\partial \theta'}\right).
\end{aligned}
$$

Assuming that

$$T^{-1}\sum_{t=1}^{T} g_t(\theta)g_t(\theta)' - T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[g_t(\theta)g_t(\theta)'\right] \xrightarrow{p} 0,$$

$$T^{-1}\sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta'} - T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right] \xrightarrow{p} 0,$$

the components of the asymptotic covariance matrix of the M-estimator W and D in Equations (A.32) and (A.33) can be consistently estimated by

$$
\begin{aligned}
\hat{W} &= T^{-1}\sum_{t=1}^{T} g_t(\hat{\theta})g_t(\hat{\theta})' \\
&= T^{-1}\sum_{t=1}^{T} 4\left[Y_t - X_t'\beta_t(\hat{\theta})\right]^2 \frac{\partial \beta_t(\hat{\theta})'}{\partial \theta}X_t X_t'\frac{\partial \beta_t(\hat{\theta})}{\partial \theta'},
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{D} &= T^{-1}\sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta'} \\
&= T^{-1}\sum_{t=1}^{T} 2\frac{\partial \beta_t(\hat{\theta})'}{\partial \theta}X_t X_t'\frac{\partial \beta_t(\hat{\theta})}{\partial \theta'}.
\end{aligned}
$$

These estimates require to compute the particular gradients $\partial \beta_t(\hat{\theta})/\partial \theta$.

A.5. Alternative GAS specifications and RLS and GAS combined with nonlinear shrinkage

In this section, several alternative specifications of the RLS- and GAS-GMVP models are provided which we developed in the course of the work on this project. First, a set of informative initializations for the recursions is presented which is based on estimates for the unconditional covariance of the return data. Second, a more extensive list of GAS specifications with differing degree of sparsity is discussed, covering – except for small modifications – the diagonal GAS employed in the final version of the paper as a particular case. Finally, we illustrate the empirical performance of these models via in- and out-of-sample comparisons to the benchmarks considered. Differences compared to the specifications in the final version of the paper are pointed out at the appropriate place. Tables and Figures regarding these alternative specifications are marked by the note '(alt)' in the captions.

A.5.1. Informative initializations

In our empirical applications, the GMVP weights have very persistent time series behavior. Hence the selection of the initial conditions β_1 and Ω_0 in the predictive recursion can become critical for the out-of-sample forecast performance when the length of the estimation window T is small. Without taking shrinkage considerations into account, if an informative initial value is to be used, a natural choice of β_1 is the OLS estimate for β in the static auxiliary regression (2.2) based on the data in the estimation period. For Ω_0 in the RLS-GMVP model we choose the corresponding sample average of $X_t X_t'$ with the coefficient γ which steers the degree of shrinkage towards the initialization (see Section 2.3.1) fixed to one; we here only consider RLS-EF. Analogously, we set $\mathbb{E}_0[X_t X_t']$ in the GMVP-GAS. We refer to these simple choices by (the suffix) *ols*. However, if the number of assets n is large in relation to T, the OLS estimate is known to be inaccurate, so that the OLS estimate provides poor starting values. For high-dimensional applications, we therefore consider another set of initial conditions indicated by (the suffix) *shr* that obtain from the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2015).

A.5.2. Restrictions for generalized autoregressive score models

When the number of assets is large a dynamic GMVP model requires a parsimonious parameterization in order to be tractable in practice. While the RLS models are parsimonious by construction, the GAS version requires parameter restrictions in order to be applicable in high dimensions. In this section, we present an extensive discussion of the GAS and propose several specifications of the model with different degrees of flexibility, starting from an unrestricted version and covering also a diagonal GAS model similar to the specification employed in the final version of the paper (see Section (2.3.2)) as a special case representing the most restrictive version. We only consider restrictions that do not violate the invariance

property.

We employ the GAS recursion as defined in Equation (2.21) combined with the Hessian evolving according to Equation (2.23), with smoothing parameter κ set equal to its typical value 0.94 (Callot et al., 2017). Noteworthy, in an initial explorative analysis we also implemented the GAS model using for H_t the Fisher information $\mathbb{E}_{t-1}\left[\nabla_t\nabla_t'\right]$ (Creal et al., 2013) and the outer product $\beta_t\beta_t'$ (similar to Opschoor et al. (2018)). However, these choices did not improve upon the trade-off between numerical stability and computational speed achieved by using the Hessian predicted by the EWMA. For other common alternatives discussed in Creal et al. (2013), including $H_t = I_n$ and $H_t = \mathbb{E}_{t-1}\left[\nabla_t\nabla_t'\right]^\nu$, $\nu \in (0,1)$ the resulting GAS model fails to be invariant w.r.t. the choice of the baseline asset (see Appendix A.2).

A necessary condition for covariance stationarity of β_t under the GAS recursion (2.21) is that the roots of B lie inside the unit circle, in which case the stationary mean of β_t obtains as

$$m \equiv \mathbb{E}(\beta_t) = (I_n - B)^{-1}c. \tag{A.34}$$

A reparametrization of the GAS recursion (2.21) in terms of this stationary mean will be instrumental for imposing parameter restrictions and for parameter estimation subject to targeting constraints. The GAS-GMVP model as given by Equations (2.21) and (2.23), when used without any restrictions on the parameters (c, B, A), is invariant w.r.t. the choice of the baseline asset, as by Lemma 3, it follows that the parameters $(c, B, A, \beta_1, \mathbb{E}_0[X_1 X_1'])$ associated with baseline asset n are in one-to-one correspondence to the parameters $(\tilde{c}, \tilde{B}, \tilde{A}, \tilde{\beta}_1, \mathbb{E}_0[\tilde{X}_1 \tilde{X}_1'])^2$ associated with baseline asset k (see Equations (A.20)-(A.24)). However, restrictions on the GAS parameters can, depending on the form of the restriction, violate the necessary one-to-one correspondence of the parameterizations. This is the case for the natural restriction to assume that the matrices $A = (a_{ij})$ and/or $B = (b_{ij})$ are diagonal, say

$$A = \text{diag}(a_{00}, a_{11}, \ldots, a_{n-1n-1}), \quad B = \text{diag}(b_{00}, b_{11}, \ldots, b_{n-1n-1}).$$

Then it is easy to verify by using (A.21) and (A.22) that $\tilde{A}$ and $\tilde{B}$ are in contrast to A and B not diagonal, unless we restrict the diagonal elements such that

$$A = \text{diag}(a_{00}, a_{11}, \ldots, a_{11}), \quad B = \text{diag}(b_{00}, b_{11}, \ldots, b_{11}), \tag{A.35}$$

or

$$A = \text{diag}(a_{00}, a_{00}, \ldots, a_{00}), \quad B = \text{diag}(b_{00}, b_{00}, \ldots, b_{00}).$$

The diagonal GAS models we use in our empirical applications are based on the restriction given in Equation (A.35) and are therefore invariant w.r.t. choice of the baseline asset.

---

[2]The parameter $\kappa = \tilde{\kappa}$ is fixed in all considered specifications of this section.

A key challenge in the GAS model is hence to derive restrictions on θ that are empirically useful (that is, achieving a good trade-off between parsimony and flexibility) without compromising the model's invariance properties. The vector of parameters in the unrestricted GAS-GMVP specification consists of $n + 2n^2$ parameters, making it difficult to estimate and prone to in-sample overfitting when the number of assets (n) is large. At the same time, the optimal degree of model complexity (simplicity) is primarily an empirical question. We therefore consider several restricted versions of the GAS model. For describing these restrictions we partition the vector of the scaled score $s_t = H_t^{-1} \nabla_t$ conformably with the vector of the GMVP mean return and weights $\beta_t = (\beta_{0t}, \beta'_{1:n-1t})'$ into $s_t = (s_{0t}, s'_{1:n-1t})'$, so that the system of GAS Equation (2.21) can be written as

$$
\begin{pmatrix} \beta_{0t+1} \\ \beta_{1:n-1t+1} \end{pmatrix} = \begin{pmatrix} c_0 \\ c_{1:n-1} \end{pmatrix} + \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & B_{11} \end{pmatrix} \begin{pmatrix} \beta_{0t} \\ \beta_{1:n-1t} \end{pmatrix} \qquad (\text{A.36})
$$
$$
+ \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & A_{11} \end{pmatrix} \begin{pmatrix} s_{0t} \\ s_{1:n-1t} \end{pmatrix}.
$$

First, we consider diagonal matrices A and B by imposing the following restrictions in Equation (A.36):

$$
b_{01} = 0, \quad b_{10} = 0, \quad B_{11} = b_{11} I_{n-1}, \quad a_{01} = 0, \quad a_{10} = 0, \quad A_{11} = a_{11} I_{n-1},
$$

where b_{11} and a_{11} are scalar parameters so that the number of parameters in A and B is reduced from $2n^2$ to 4. Under this diagonal restriction the dynamic structure of the GMVP mean return $\beta_{0,t+1}$ (directed by the parameters b_{00} and a_{00}) is allowed to differ from that of the GMVP weights $\beta_{i,t+1}, i = 1, \dots, n-1$ (directed by the parameters b_{11} and a_{11}), but the dynamic structure for the weights is restricted to be the same across all assets. The latter is needed to ensure that the restricted diagonal model remains to be invariant w.r.t. the choice of baseline asset, which does not allow the elements in the diagonal matrices B_{11} and A_{11} to differ (see Appendix A.2).

Second, we consider the restriction that the expected GMVP return $\beta_{0,t+1}$ is constant over time which obtains by using in Equation (A.36)

$$
b_{00} = 0, \quad b_{01} = 0, \quad a_{00} = 0, \quad a_{01} = 0,
$$

so that the number of parameters is reduced by $2n$. This restriction appears to be reasonable since portfolio returns are typically difficult to predict based on past information, at least at the daily frequency (Cochrane, 2005, Chapter 20).

Third, we consider the restriction that the long-run mean of the portfolio weights correspond to the equally weighted portfolio, so that $\mathbb{E}(\beta_{it}) = 1/n$ for $i = 1, \dots, n-1$. Using Equation (A.34) and partitioning $m = \mathbb{E}(\beta_t)$ conformably with the vector of intercepts

$c = (c_0, c'_{1:n-1})'$, this restriction can be represented as

$$\begin{pmatrix} c_0 \\ c_{1:n-1} \end{pmatrix} = \left[I_n - \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & B_{11} \end{pmatrix} \right] \begin{pmatrix} m_0 \\ m_{1:n-1} \end{pmatrix}, \quad \text{with} \quad m_{1:n-1} = \iota_{n-1}/n. \qquad (A.37)$$

By this restriction, fixing $n-1$ parameters, all GMVP weights are forced to fluctuate around the benchmark value $1/n$ defined by the equally weighted portfolio. Our use of this benchmark follows DeMiguel et al. (2009); Candelon et al. (2012); Frey and Pohlmeier (2016) who consider shrinkage of the GMVP weights towards equality in a static framework. More broadly, the popularity of the equally weighted portfolio can perhaps be explained by its simplicity and the fact that it avoids estimation errors (DeMiguel et al., 2007).

Finally, we consider a reduction of the number of parameters by using a targeting estimation approach that replaces the long-run mean vector of the GMVP weights m in Equation (A.37) by a sample estimate. A simple estimate obtains by running the static auxiliary regression of Kempf and Memmel (2006) using OLS or, equivalently, by replacing in Equation (2.1) the population covariance Σ by the sample covariance matrix. Moreover, to robustify the targeting approach against large dimensions, we also consider a specification using the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2015) to estimate Σ.

The GAS models we consider in our empirical work include the unrestricted model and models with the four restrictions described above, as well as combinations thereof. Table A.3 lists the resulting model specifications: The (non-diagonal) model without any restriction (unr.-GAS), the diagonal model (d-GAS), the diagonal model with constant mean return (d-GAS-cβ_0), equal long-run weights (d-GAS-ew), constant mean return combined with equal long-run weights (d-GAS-cβ_0-ew), constant mean return combined with targeting towards the OLS estimates of the long-run weights (d-GAS-cβ_0-ols-ta) and combined with targeting towards the nonlinear shrinkage estimates of the long-run weights (d-GAS-cβ_0-shr-ta). In accordance with how its target is constructed, we use *shr* initial conditions for the d-GAS-cβ_0-shr-ta model. For all other GAS specifications, the *ols* initializations are used (see Section A.5.1.

A.5.3. Analytical gradients

For inferring the asymptotic distribution for the parameter estimates under the NLS optimization (see Appendix A.4) and in order to alleviate the numerically challenging optimization of the loss function, it can be helpful to provide the analytical solutions for the gradients. They are given below for the RLS-EF and GAS models employed in this section, i.e., with $\kappa = 0.94$ and $\gamma = 1$ fixed.

A.5. Alternative GAS specifications and RLS and GAS combined with nonlinear shrinkage

Table A.3.: List of GAS model specifications (alt.)

Model	Restriction	# Params
unr.-GAS	none	$2n^2 + n$
d-GAS	$B_{11} = b_{11}I_{n-1}, \quad A_{11} = a_{11}I_{n-1},$ $b_{01} = a_{01} = 0, \quad b_{10} = a_{10} = 0$	$n + 4$
d-GAS-cβ_0	& $b_{00} = a_{00} = 0$	$n + 2$
d-GAS-ew	& $m_{1:n-1} = \iota_{n-1}/n$	5
d-GAS-cβ_0-ew	& $b_{00} = a_{00} = 0, \ m_{1:n-1} = \iota_{n-1}/n$	3
d-GAS-cβ_0-ols-ta	& $b_{00} = a_{00} = 0, \ m$: OLS estimate	2
d-GAS-cβ_0-shr-ta	& $b_{00} = a_{00} = 0, \ m$: nonlinear shrinkage estimate	2

Note: List of GAS model specifications according to Equations (A.36) and (A.37); n is the number of assets in the portfolio and $m = \mathbb{E}(\beta_t)$. '# Params' indicates the number of model parameters to be estimated by the M-estimator given in Equation (2.8).

For the RLS-EF model

$$\beta_t = \beta_{t-1} + \Omega_{t-1}^{-1}X_{t-1}(Y_{t-1} - X'_{t-1}\beta_{t-1})$$
$$= \beta_{t-1} + \frac{\Omega_{t-2}^{-1}X_{t-1}}{\lambda + X'_{t-1}\Omega_{t-2}^{-1}X_{t-1}}(Y_{t-1} - X'_{t-1}\beta_{t-1}),$$

with $\theta = \lambda$ the first derivative of $\beta_t(\lambda)$ obtains recursively as

$$\frac{\partial \beta_t(\lambda)'}{\partial \lambda} = \frac{\partial \beta_{t-1}(\lambda)'}{\partial \lambda}$$
$$+ \left\{ (\Psi_{t-2}X_{t-1})'(\lambda + X'_{t-1}\Omega_{t-2}^{-1}X_{t-1}) \right.$$
$$\left. - (\Omega_{t-2}^{-1}X_{t-1})'(1 + X'_{t-1}\Psi_{t-2}X_{t-1}) \right\}$$
$$\times \frac{Y_{t-1} - X'_{t-1}\beta_{t-1}(\lambda)}{(\lambda + X'_{t-1}\Omega_{t-2}^{-1}X_{t-1})^2}$$
$$- \frac{X'_{t-1}\Omega_{t-2}^{-1}}{\lambda + X'_{t-1}\Omega_{t-2}^{-1}X_{t-1}}\frac{\partial \beta_{t-1}(\lambda)'}{\partial \lambda}X_{t-1},$$

where

$$\Psi_t = \frac{\partial \Omega_t^{-1}}{\partial \lambda} = -\Omega_t^{-1}\frac{\partial \Omega_t}{\partial \lambda}\Omega_t^{-1}.$$

For the GAS model

$$\beta_t = c + B\beta_{t-1} + AH_{t-1}^{-1}\nabla_{t-1},$$

with $\theta = (c', \text{vec}(B)', \text{vec}(A)')'$, the corresponding recursion for the first derivative of $\beta_t(\theta)$ is

$$
\begin{aligned}
\frac{\partial \beta_t(\theta)'}{\partial \theta} \quad = \quad & \frac{\partial c'}{\partial \theta} + \frac{\partial \text{vec}(B)'}{\partial \theta}(\beta_{t-1}(\theta) \otimes I_n) \\
& + \frac{\partial \text{vec}(A)'}{\partial \theta}\left[(H_{t-1}^{-1}\nabla_{t-1}) \otimes I_n\right] \\
& + \frac{\partial \beta_{t-1}(\theta)'}{\partial \theta}\left(2X_{t-1}X_{t-1}'H_{t-1}^{-1}A' + B'\right).
\end{aligned}
$$

A.5.4. Empirical results

In this section we apply our previously discussed GMVP prediction models to historical return data, and compare them to the benchmark methods that are described in Section 2.4.1. We consider both daily data and monthly data. In doing so, we intend to cover various practically relevant scenarios regarding the ratio n/T (number of assets divided by length of the time series). For the experiments with daily returns we use the same data set and design, as well as asset universe selection as employed in the final version of the paper, under the restriction that the sample covers only the time span until 2016 and a maximum of $n = 200$ cross-sectional units. Precisely, the sample covers the period from 01/02/2002 to 12/06/2016 for a total of T = 3759 trading days.

In-sample results for daily data

Estimating the heavily parameterized non-diagonal unrestricted GAS specification is prohibitively difficult for a large number of stocks. In an initial experiment, we thus consider a low-dimensional application of our proposed GMVP models to five of the 200 stocks. This experiment allows us to compare the seven GAS specifications in Table A.3 and the RLS-EF model and to investigate the impact of different degrees of sparsity of the GAS parameter matrices c, A and B. The five stocks we use for this experiment are those with the largest market capitalization at the last trading day of the estimation period: Apple, Microsoft, ExxonMobil, Amazon and Johnson & Johnson.

Table A.4.: Average in-sample GMVP loss and variance for $n = 5$ stocks (alt.)

	unr.-GAS	d-GAS	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0	d-GAS -ew	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
Avg. loss	0.928	1.029	1.055	1.055	1.055	1.050	1.071	1.067	1.067
Portf. var.	0.928	1.046	1.055	1.055	1.055	1.066	1.071	1.061	1.061

Note: Average in-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS-EF models for $n = 5$ stocks. The sample period ranges from 01/02/2002 to 12/06/2016 ($T = 3759$).

Table A.4 reports the average in-sample GMVP loss for the seven GAS specifications and the two versions of the RLS-EF model (initialized by the OLS and nonlinear shrinkage approaches). As expected, the lowest average loss is attained by the unrestricted (non-diagonal) GAS specification with 55 parameters, which allows for the largest degree of flexibility in approximating the dynamic behavior of the GMVP weights. The models with the largest average losses are the d-GAS-cβ_0-ew (constant expected portfolio returns and equal long-run means of the portfolio weights) with three parameters and the RLS-EF with only one parameter. However, even for a dimension as low as in the present experiment ($n = 5$), NLS based parameter estimation in the unrestricted GAS model turned out to be numerically challenging[3]. Moreover, the parameter estimates for the unrestricted GAS model (not reported here) reveal that none of the model's 55 parameters is significantly different from zero at the 10% level, indicating that the model is clearly over-parameterized. With regard to out-of-sample forecast accuracy (analyzed below), these results suggest using restrictions to eliminate many of the unnecessary parameters. By contrast, the estimates of the d-GAS parameter b_{11} are highly significant and exceed 0.98 for all diagonal GAS specifications, indicating high temporal persistence of the GMVP weights. The estimates for a_{00} and b_{00} are for all specifications insignificant at conventional levels, so that there is no evidence for predictable dynamics in the conditional mean of the daily GMVP return. The estimate for the forgetting factor of the RLS-EF model λ is statistically highly significant and its estimated value is 0.988, which implies that observations from a year ago still have a weight of 5% in the current GMVP prediction, whereas observations from two years ago have a weight of 0.2%.
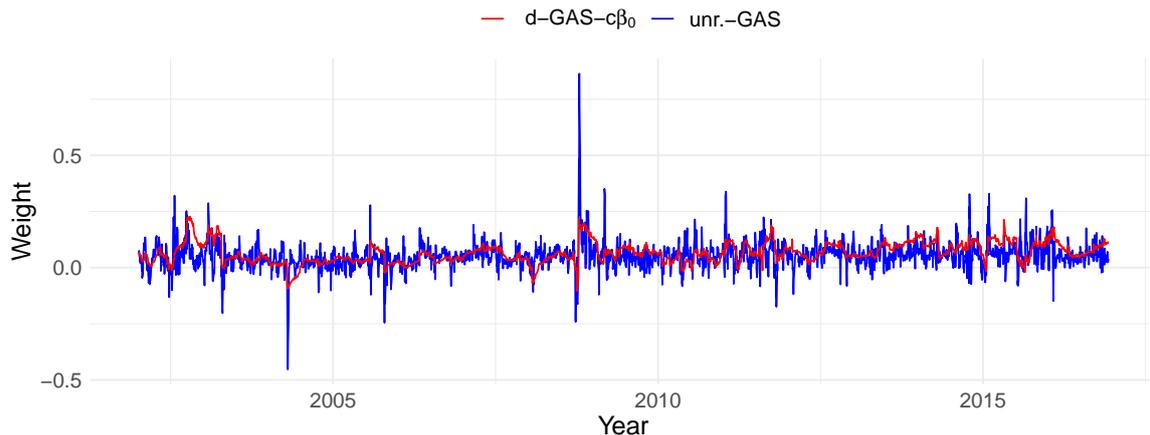
Figure A.6.: Estimated GMVP weights for the Johnson & Johnson stock in a portfolio of $n = 5$ assets, for the (fully unrestricted) unr.-GAS (blue line) and the d-GAS-cβ_0 model (red line).

In addition to these results we find that the predictions of the portfolio weights obtained under the sparsely parameterized diagonal GAS specifications and the RLS-EF model are

---

[3]One reason of the numerical problems appears to be the large number of local minima of the average GMVP loss function.

less noisy than those for the unrestricted GAS model, so that the former ask for less portfolio regrouping. This is illustrated in Figure A.6, which shows the time series plots of the portfolio weights for the Johnson & Johnson stock predicted under the unrestricted GAS and the d-GAS-cβ_0. While the trends of the two series are very similar, the weight predictions of d-GAS-cβ_0 are less volatile than those of the unrestricted GAS. In particular, d-GAS-cβ_0 does not feature the economically questionable jumps predicted by the unrestricted GAS. In addition to the average GMVP loss, we report in Table A.4 the sample variance of the predicted GMVP returns. The comparison of the values for the two performance measures shows that they are typically close to each other. This suggests that the estimated GMVP mean returns β_{0t} are fairly stable over time, so that the observed small differences in the two measures are in line with the assumption on the lack of predictability of GMVP returns.

Based on the results of our low-dimensional application, it seems empirically and practically reasonable to focus on the sparsely parameterized versions of our proposed GMVP models, i.e., the diagonal GAS models and the RLS-EF model. The average in-sample loss and sample variance of the portfolio returns for those models applied to all $n = 200$ stocks in our daily data set are reported in Table A.5 which also provides results for the benchmark models described in Section 2.4.1.

Table A.5.: Average in-sample GMVP loss and variance for $n = 200$ stocks (alt.)

	d-GAS	d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -ew	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
Avg. loss	0.350	0.351	0.354	0.355	0.354	0.355	0.352	0.352
Portf. var.	0.350	0.351	0.354	0.355	0.355	0.355	0.352	0.352
	DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve		
Avg. loss	0.425	0.425	0.355	0.356	0.357	1.603		
Portf. var.	0.425	0.425	0.355	0.355	0.357	1.601		

Note: Average in-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS-EF models and the benchmark models for $n = 200$ stocks. The sample period ranges from 01/02/2002 to 12/06/2016 ($T = 3759$).

Table A.5 shows that the RLS-EF model and nearly all d-GAS models attain smaller in-sample loss than the competing benchmark models. For example, the two-parameter d-GAS-cβ_0-ols-ta and d-GAS-cβ_0-shr-ta model, and even the single parameter RLS-EF approach, result in portfolio predictions with much lower in-sample losses than the portfolios predicted by the highly parameterized DCC with 602 parameters. It is only the static OLS estimator which is on par with the d-GAS-cβ_0-shr-ta and the d-GAS-cβ_0-eω model. Next, we find that the increase in the average loss resulting from imposing the cβ_0-restriction of constant expected portfolio returns in the d-GAS model is negligible and is only about 0.3%. The loss increase for the ew-restriction of equal long-run means of the portfolio weights is somewhat

larger at 1%. This suggests that the stationary equilibrium differs from the naïve equally weighted portfolio which is consistent with the result that the naïve portfolio model produces by far the largest average loss among all models. It is also in line with our finding that the optimal degree of shrinkage towards the naïve portfolio for the shrinkage estimators is close to zero, so that the weight predictions and the resulting average losses for the static OLS, SHR-l and SHR-nl estimators are very close to each other.

As in the low-dimensional application, the parameter estimates of the GAS models (not reported here) indicate high persistence of the GMVP portfolio weights with estimates for b_{11} larger than 0.9 for all d-GAS specifications. The estimate for the forgetting factor λ in the RLS-EF model is 0.999, which is also close to the value found in the low-dimensional experiment.

All in all, the results of the in-sample experiment show that all considered loss function based dynamic GMVP models perform broadly similar and exhibit a better in-sample fit compared to the DCC model fitted by a likelihood based estimation technique and then used within the common plug-in approach.

Out-of-sample results for daily data

The out-of-sample period starts on 01/03/2007 and ends on 12/06/2016 which results in a total of 2,501 daily point forecasts. We consider estimation window lengths of $T = 250$ (one year) and $T = 1250$ (five years) and portfolio sizes of $n = 50$ and $n = 200$.

Since the in-sample analysis yields virtually no evidence for temporal variation in the daily expected portfolio returns, we focus on the d-GAS specifications with the $c\beta_0$-restriction in our out-of-sample analysis. Table A.6 provides the average out-of-sample loss and portfolio variance of these d-GAS-$c\beta_0$ models, the RLS-EF and the benchmark models. The top panel reports the results for the portfolio size $n = 50$ based on the 1-year and the 5-year estimation window with a ratio n/T of 0.2 and 0.04, respectively, and the bottom panel the corresponding results for $n = 200$ with value for n/T of 0.8 and 0.16. For assessing the statistical significance of differences in the average out-of-sample loss across models, we use the model confidence set (MCS) approach of Hansen et al. (2011) (for details see Section 2.4.2).

Table A.6 reveals that, for all four n/T-scenarios, the two best performing specifications within our proposed GMVP approach are the d-GAS-$c\beta_0$ model with targeting based on the nonlinear shrinkage approach (d-GAS-$c\beta_0$-shr-ta) and RLS-EF initialized by the nonlinear shrinkage approach (RLS-EF-shr). We also see that their performance gains relative to RLS-EF-ols and the other d-GAS-$c\beta_0$ models consistently increase with the ratio n/T. This finding indicates that a precise estimator of the long-run-weights (provided by nonlinear shrinkage) is particularly valuable when the number of assets is large relative to the sample size. The finding is also fully in line with the forecast improvements we find when moving from standard DCC with OLS targeting (DCC) to DCC with nonlinear shrinkage targeting

Table A.6.: Average out-of-sample GMVP loss and variance (alt.)

				$n = 50$			
$T = 1250$ $n/T = 0.04$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
	Avg. loss	0.671	0.672	0.665	0.675	0.639	**0.630**
	Portf. var.	0.671	0.672	0.665	0.675	0.638	0.630
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.654	0.647	0.669	0.667	0.665	1.574
	Portf. var.	0.654	0.647	0.669	0.667	0.665	1.574
$T = 250$ $n/T = 0.20$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
	Avg. loss	1.778	0.832	0.696	0.833	0.751	**0.654**
	Portf. var.	1.764	0.829	0.693	0.830	0.748	0.651
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.737	0.673	0.755	0.729	0.677	1.584
	Portf. var.	0.736	0.672	0.752	0.729	0.677	1.585
				$n = 200$			
$T = 1250$ $n/T = 0.16$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
	Avg. loss	0.625	0.526	0.506	0.525	0.510	0.492
	Portf. var.	0.622	0.526	0.506	0.525	0.510	0.491
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.493	**0.474**	0.524	0.521	0.504	1.808
	Portf. var.	0.493	0.474	0.524	0.521	0.504	1.808
$T = 250$ $n/T = 0.80$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS-EF -ols	RLS-EF -shr
	Avg. loss	1.767	2.666	0.513	4.387	1.543	**0.486**
	Portf. var.	1.767	2.661	0.511	4.383	1.538	0.484
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.984	0.510	1.575	0.864	0.495	1.813
	Portf. var.	0.984	0.510	1.571	0.864	0.495	1.813

Note: Average out-of-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS-EF models and the benchmark models. Parameter estimation is based on a sample of length T. The out-of-sample period ranges from 01/03/2007 to 12/06/2016 (2,501 observations). Bold numbers indicate the smallest average GMVP loss and grey cells indicate that the model belongs to the 90% Model Confidence Set.

(DCC-nl) and from static OLS to the nonlinear shrinkage estimator (SHR-nl). The comparison of all competing models shows that RLS-EF-shr belongs to the 90% MCS for all n/T

ratios and also has the smallest average loss, except in the scenario with $(n, T) = (200, 1250)$, where DCC-nl performs best. Finally, we find that the performance of the static SHR-nl estimator improves compared to the dynamic models when the sample size T decreases, which is to be expected, since the shorter the (rolling) estimation window, the easier it is for a static approach to adapt to local parameter change.

Out-of-sample results for monthly data

In this section, we analyze the robustness of our previous results to changes in the sampling frequency by considering the out-of-sample forecasting performance of the GMVP models when applied to monthly asset returns. For this experiment we rely upon four data sets used by DeMiguel et al. (2009), which we have extended to cover the period from 07/1963 to 02/2019. They consist of returns for Fama-French portfolios sorted by size and book-to-market ratio and industry portfolios representing the U.S. stock market: The first data set contains six Fama-French portfolios (6-FF), the second one 25 Fama-French portfolios (25-FF), the third one 10 industry portfolios (10-Ind), and the last one 48 industry portfolios (48-Ind)[4]. As in our application to daily returns, we consider one-period-ahead forecasts and use a rolling window for parameter estimation. The window length is set equal to $T = 120$ (10 years) and the out-of-sample period ranges from 07/1973 to 02/2019, which yields a total of 548 monthly forecasts for each data set.

Table A.7 contains the resulting average out-of-sample GMVP loss of our proposed models and the competing alternatives. They show that for all four portfolios, the best performing model in the class of RLS-EF and d-GAS models attains smaller average loss than both versions of the DCC approach. For the 6-FF, 25-FF and 10-Ind portfolios, for which the n/T-ratio is at most 0.21, the static SHR-l estimator attains the smallest loss. However, the RLS-EF-shr model is always contained in the 90% MCS, indicating that the loss difference between RLS-EF-shr and SHR-l is not statistically significant. As explained above, one possible explanation for the relatively good performance of this static SHR estimator is that the estimation window in this experiment is rather short ($T = 120$), much shorter than in our previous experiments with daily data. For the 48-Ind portfolio for which we have the most challenging scenario in terms of the n/T-ratio, the best model is the RLS-EF-shr with a performance which is significantly better than that for all competing models. Furthermore, we observe that RLS-EF-shr belongs to the 90% MCS for all four portfolios, which is fully in line with the results for the experiments based on daily data.

Superior Predictive Ability Test Results

As a robustness check for the MCS results for our out-of-sample forecasting experiments, Table A.8 reports p-values of the test for Superior Predictive Ability (SPA) (see Section

---

[4]The data has been obtained from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Table A.7.: Average out-of-sample GMVP loss for monthly portfolio returns (alt.)

	6-FF	25-FF	10-Ind	48-Ind
n/T	0.05	0.21	0.08	0.49
d-GAS-cβ_0	17.14	19.81	14.03	21.35
d-GAS-cβ_0-ols-ta	16.41	15.90	13.69	19.79
d-GAS-cβ_0-shr-ta	16.54	13.80	13.22	13.96
d-GAS-cβ_0-ew	16.06	15.75	13.66	20.49
RLS-EF-ols	17.60	13.90	14.04	10.10
RLS-EF-shr	17.60	13.67	14.04	**10.07**
DCC	17.22	16.99	13.77	16.35
DCC-nl	17.16	16.16	13.45	13.22
OLS	15.72	14.83	12.86	17.30
SHR-l	**15.54**	**13.27**	**12.63**	14.84
SHR-nl	15.72	13.63	12.79	13.11
naïve	24.49	26.48	18.35	23.79

Note: Average out-of-sample GMVP loss of the diagonal GAS models, the RLS-EF and the benchmark models for monthly portfolio returns. The length of the estimation window is 10 years ($T = 120$), the out-of-sample forecasting period ranges from 07/1973 to 02/2019 (548 observations). Bold numbers indicate the smallest average GMVP loss and grey shaded cells indicate that the model belongs to the 90% Model Confidence Set.

A.3.3). The results in Table A.8 are qualitatively very similar to the results based on the MCS approach. In particular, successful models that are included in the MCS typically achieve high SPA *p*-values when used as a benchmark model. Conversely, for poor performing models that are not included in the MCS, the SPA *p*-value is typically very small.

Table A.8.: Hansen (2005) test for superior predictive ability results (alt.)

	Daily data				Monthly data			
					6-FF	10-Ind	25-FF	48-Ind
n	50	50	200	200	6	10	25	48
T	250	1250	250	1250	120	120	120	120
d-GAS-cβ_0	0.03	0.00	0.00	0.00	0.06	0.02	0.00	0.00
d-GAS-cβ_0-ols-ta	0.00	0.00	0.00	0.00	0.19	0.02	0.00	0.00
d-GAS-cβ_0-shr-ta	0.03	0.01	0.30	0.19	0.14	0.21	0.55	0.00
d-GAS-$c\beta_0$-ew	0.00	0.01	0.01	0.00	0.64	0.05	0.00	0.00
RLS-EF-ols	0.00	0.64	0.00	0.11	0.07	0.14	0.00	0.01
RLS-EF-shr	0.99	0.99	0.97	0.75	0.08	0.14	0.82	1.00
DCC	0.00	0.00	0.00	0.00	0.11	0.01	0.00	0.00
DCC-nl	0.75	0.70	0.80	0.98	0.11	0.35	0.02	0.00
OLS	0.00	0.00	0.00	0.00	0.89	0.54	0.00	0.00
SHR-l	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00
SHR-nl	0.62	0.01	0.96	0.40	0.79	0.11	0.24	0.00
naïve	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: p-values of the Hansen (2005) test for superior predictive ability. Rows correspond to benchmark models, columns correspond to forecast experiments. The one-sided Null hypothesis is that the benchmark is at least as good as all competitors.

Appendix B.

Appendix for Chapter 3

B.1. Priors and hyperparameters

Following Bitto and Frühwirth-Schnatter (2019), we impose $p_{0,i} \sim \mathcal{IG}(\nu_p, (\nu_p-1)c_p)$, $\nu_p = 20$, $c_p = 1$, for $i = 0, \ldots, n$ and for the shrinkage priors $d_{01} = d_{02} = e_{01} = e_{02} = 0.001$, as well as $b^\tau = 1$, $c^\tau = 10$ with the proposal for $\log(a^\tau) \sim \mathcal{N}(0,1)$, as this applies only for the double Gamma prior. For the weights in the *no shr* specification and for all specifications for the conditional expectation, the priors of level and time-variation are set in an uninformative, but also zero-centered way, i.e., $\mathcal{N}(0,10)$. The volatility priors are given follows: $\sigma^2 \mid C_0 \sim \mathcal{IG}(c_0, C_0)$, $C_0 \sim \mathcal{G}(g_0, G_0)$, where $c_0 = 2.5$, $g_0 = 2.5$, $G_0 = g_0/\mathbb{V}(\widehat{y - X\hat{\beta}}_{OLS}) \times (c_0 - 1)$ for the model with homoscedastic errors and in the SV model $\mu_h \sim \mathcal{N}(0,1)$, $\sigma_h^2 \sim \mathcal{IG}(0.5, 0.5)$, and $(\phi_h + 1)/2 \sim \mathcal{B}_1(100, 4)$, which implies $\mathbb{E}(\phi_h) = 0.992$ and $\mathbb{V}(\phi_h) = 0.0002$.

B.2. Further derivations

B.2.1. MCMC sampling algorithm

We utilize MCMC methods for Bayesian posterior analysis and use the Gibbs approach to simulate from the joint posterior of the parameters and the latent states given by

$$
\begin{aligned}
\pi(\gamma_{0:T}, & h_{0:T}, \alpha, \sqrt{q}, \tau^2, \xi^2, \lambda^2, \kappa^2, \phi_h, \mu_h, \sigma_h^2, a^\tau, p_0 \mid y_{1:T}) \\
& \propto f_\theta(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \ f(\gamma_{1:T}) \ f_\theta(h_{0:T}) \ p(\gamma_0) \\
& \quad \times p(\alpha, \sqrt{q} \mid \tau^2, \xi^2) \ p(\tau^2, \xi^2 \mid \lambda^2, \kappa^2) \ p(\lambda^2) \ p(\kappa^2) \ p(a^\tau) \\
& \quad \times p(\mu_h, \phi_h, \sigma_h^2) \times p(p_0)
\end{aligned}
$$

for the TVP-GMVP-SV-*shr* model. In the alternative SV specifications with a random walk for the log volatility process (see Appendix B.3.1), the volatility parameters $(\mu_h, \sigma_h^2, \phi_h)$ are replaced by R, and in the specification with the Uhlig (1994, 1997) volatility process (see Appendix B.3.2), they are replaced by ν_h and λ_h. In the model without stochastic volatility, the states h and the associated parameters of the volatility process are replaced by a time-constant volatility σ^2 which is equipped with a hierarchical Gamma prior (see Section

B.1). Exploiting the conjugacy to the Gaussian likelihood, conditional posteriors are mostly available in closed form. In the *no shr* specifications as well as for the parameters associated to the first element of β_t corresponding to the conditional expectation of the GMVP, Steps 3) and 4) do not apply.

The implementation of the Gibbs sampler for the TVP-GMVP-SV-*shr* model is briefly described in the following. Note that Steps 1) − 4) and Step 6) of the algorithm are adapted from Bitto and Frühwirth-Schnatter (2019), Algorithm 1.

1) Updating $\gamma_{0:T}$:

Choosing a Gaussian prior $f(\gamma_{1:T}) \times p(\gamma_0)$ yields a linear Gaussian SSM for $\gamma_{0:T}$ (for given $\theta, h_{1:T}$). To draw from full conditional posterior of $\gamma_{0:T}$, that is,

$$\pi(\gamma_{0:T} \mid \theta, y_{1:T}, h_{0:T}) \propto f_\theta(y_{1:T} \mid h_{1:T}, \gamma_{1:T}) \, f(\gamma_{1:T}) \, p(\gamma_0),$$

one can apply the method of Forward Filtering Backward Sampling (FFBS; Carter and Kohn, 1994 and Frühwirth-Schnatter, 1994) or the computationally beneficial *all without a loop* sampler (AWOL; Rue, 2001 and McCausland et al., 2011). Depending on the RAM of the computer, one may rather apply FFBS when the dimensions of the system (i.e., number of time periods and number of parameters, i.e., assets here) are very large.

2) Updating α and $\sqrt{q}$

Using multivariate normal conjugate priors, i.e., for $i = 0, \ldots, n$,

$$\alpha_i \mid \tau_i^2 \sim \mathcal{N}(0, \tau_i^2), \qquad\qquad \sqrt{q}_i \mid \xi_i^2 \sim \mathcal{N}(0, \xi_i^2),$$

the full conditional posterior

$$\pi(\alpha, \sqrt{q} \mid \theta_{-(\alpha,\sqrt{q})}, \gamma_{0:T}, y_{1:T}, h_{0:T}) \propto f_\theta(y_{1:T} \mid h_{1:T}, \gamma_{1:T}) \, p(\alpha, \sqrt{q} \mid \tau^2, \xi^2)$$

directly obtains.

In order to increase the sampling efficiency, we apply the interweaving step from Bitto and Frühwirth-Schnatter (2019) which builds on redrawing the parameters α and $\sqrt{q}$ in the centered parameterization of the model (see Equations (3.7) and (3.8)). Note that the posterior density of the scale parameters $\sqrt{q}$ is symmetric around zero by definition. Thus, if for some cross-sectional unit i the variance $\sqrt{q}_i^2$ is different from zero, then the posterior density of $\sqrt{q}_i$ is likely to be bimodal and if we find that the posterior density of $\sqrt{q}$ is unimodal, then the unknown variance is likely to be zero.

3) Updating τ^2 and ξ^2, and λ^2 and κ^2:

Based on the hierarchical structure of the priors, i.e., for $i = 1, \ldots, n$:

$$\tau_i^2 \mid \lambda^2 \sim \mathcal{G}(a^\tau, a^\tau \lambda^2), \qquad\qquad \lambda^2 \sim \mathcal{G}(d_{01}, d_{02}),$$
$$\xi_i^2 \mid \kappa^2 \sim \mathcal{G}(a^\xi, a^\xi \kappa^2), \qquad\qquad \kappa^2 \sim \mathcal{G}(e_{01}, e_{02}),$$

where d_{01}, d_{02} and e_{01}, e_{02} are fixed a-priori, we block the conditional posterior as follows

$$\begin{aligned} \pi(\tau^2, \xi^2, &\lambda^2, \kappa^2 \mid \theta_{-(\tau^2, \xi^2, \lambda^2, \kappa^2)}, \gamma_{0:T}, y_{1:T}, h_{0:T}) \\ &\propto p(\tau^2, \xi^2, \lambda^2, \kappa^2 \mid \alpha, \sqrt{q}) \, p(\tau^2, \xi^2 \mid \lambda^2, \kappa^2) \, p(\lambda^2, \kappa^2 \mid \tau^2, \xi^2). \end{aligned}$$

This implies a Generalized inverse Gamma posterior distribution for τ_i^2, ξ_i^2, $i = 1, \ldots, n$, and the conditional posteriors for λ^2 and κ^2 directly obtain from the Gamma conjugate priors.

4) Updating a^τ:

In the shrinkage specification, we impose $\mathcal{G}(b^\tau, c^\tau)$ distributions for a^τ ($a^\xi = 1$ is fixed). Draws from the conditional posterior

$$\pi(a^\tau \mid \theta_{-(a^\tau)}, \gamma_{0:T}, y_{1:T}, h_{0:T}) \propto p(a^\tau \mid \lambda^2, \tau^2)$$

are obtained by sampling a^τ using a RW-MH step based on proposing $\log(a^{\tau, new}) \sim \mathcal{N}(\log a^\tau, 1)$.

5) Updating $h_{0:T}$ and $(\mu_h, \sigma_h^2, \phi_h)$:

In order to draw the volatility states and parameters, we use the standard approach of Kim et al. (1998), replacing the normal approximation by the 10-point mixture of Omori et al. (2007), which yields a measurement equation that is approximately normal, and hence, results in an approximately Gaussian state-space model for the volatility states. Therefore, the derivation of the conditional posterior

$$\begin{aligned} \pi(h_{0:T}, \mu_h, &\sigma_h^2, \phi_h \mid \theta_{-(\mu_h, \sigma_h^2, \phi_h)}, \gamma_{0:T}, y_{1:T}) \\ &\propto f_\theta(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \, p(h_{0:T} \mid \mu_h, \sigma_h^2, \phi_h) \, p(\mu_h, \sigma_h^2, \phi_h) \end{aligned}$$

is analogue to Step 1) and h can be drawn using FFBS or the AWOL sampler. For $(\mu_h, \sigma_h^2, \phi_h)$, we impose $p(\mu_h, \sigma_h^2, \phi_h) = p(\mu_h)p(\sigma_h^2)p(\phi_h)$. A rather uninformative Gaussian prior is imposed for μ_h, a Beta prior for $((\phi + 1)/2)$ and a conjugate inverse Gamma prior for σ_h^2.

Alternative (i): *5) "Random walk for* $\log(\sigma_t^2)$*" Updating* $h_{0:T}$ *and* R:

In order to draw the volatility states, we use the standard approach of Kim et al. (1998)[1] which yields a measurement equation that is approximately normal, and hence, results in an approximately Gaussian state-space model for the volatility states. Hence the derivation of the conditional posterior

$$\pi\left(h_{0:T}, R \mid \theta_{-(R)}, \gamma_{0:T}, y_{1:T}\right) \propto f_{\theta}(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \, p(h_{1:T} \mid R) \, p(R, h_0)$$

is analogue to Step 1) and h can be drawn using FFBS or the AWOL sampler. For R, we select an independent Inverse Gamma conjugate prior such that we can directly draw from its conditional posterior (Inverse Gamma).

Alternative (ii): *5) "Beta process for* $1/\sigma_t^2$*" Updating* $h_{1:T}$ *and* ν_h *(and* λ_h*):

For sampling $h_{1:T}$ and ν_h, λ_h from their joint conditional posterior

$$\pi(h_{1:T}, \nu_h, \lambda_h \mid \theta_{-(\nu_h, \lambda_h)}, \gamma_{0:T}, y_{1:T}) \propto f_{\theta}(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \, p(h_{1:T}) \, p(\nu_h, \lambda_h),$$

ν_h and λ_h are simulated marginally of $h_{1:T}$ from

$$
\begin{aligned}
\pi(\nu_h, \lambda_h \mid & \theta_{-(\nu_h, \lambda_h)}, \gamma_{0:T}, y_{1:T}) \\
& \propto \left[\int f_{\theta}(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \, f_{\theta}\left(h_{1:T}\right) \mathrm{d}h_{1:T}\right] \, p(\nu_h, \lambda_h) \\
& = f_{\theta}(y_{1:T} \mid \gamma_{1:T}) \, p(\nu_h, \lambda_h).
\end{aligned}
$$

Here, we make use of the closed-form available integrated likelihood $f_{\theta}(y_{1:T} \mid \gamma_{1:T})$ to simulate ν_h, λ_h jointly by a standard Gaussian RW-MH. Factorizing the integrated likelihood yields

$$
\begin{aligned}
f_{\theta}(y_{1:T} \mid \gamma_{1:T}) &= \prod_{t=1}^{T} f_{\theta}(y_t \mid \gamma_{1:t}, y_{0:t-1}), \\
f_{\theta}(y_t \mid \gamma_{1:t}, y_{0:t-1}) &= \int f_{\theta}(y_t \mid h_t, \gamma_{1:t}) \, f_{\theta}(h_t \mid y_{0:t-1}) \, \mathrm{d}h_t,
\end{aligned}
$$

---

[1]We again replace their normal approximation by the 10-point mixture of Omori et al. (2007).

where $f_\theta(y_t \mid h_t, \gamma_{1:t}) = f_\theta(y_t \mid h_t, \gamma_t)$. Inserting $(y_t \mid h_t, \gamma_t) \sim \mathcal{N}(x'_t\beta_t, 1/h_t)$ and from Windle and Carvalho (2014, Proposition 1), $(h_t \mid \theta, y_{0:t-1}) \sim \mathcal{G}(\nu_h/2, \sigma^2_{h,t-1}\lambda_h/2)$, we obtain

$$\int f_\theta(y_t \mid h_t, \gamma_{1:t}) \ f_\theta(h_t \mid y_{0:t-1}) \ \mathrm{d}h_t$$

$$= \int \frac{h_t^{1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\varepsilon_t^2 h_t\right\} \frac{(\sigma^2_{h,t-1},\lambda_h)^{\nu_h/2} 2^{-\nu_h/2}}{\Gamma(\nu_h/2)} h_t^{\nu_h/2-1} \exp\left\{-\frac{1}{2}(\sigma^2_{h,t-1}\lambda_h)h_t\right\} \ \mathrm{d}h_t$$

$$= \frac{2^{-\frac{1+\nu_h}{2}}(\sigma^2_{h,t-1},\lambda_h)^{\nu_h/2}}{\Gamma(\nu_h/2)\pi^{1/2}} \int h_t^{\frac{\nu_h-1}{2}} \exp\left\{-\frac{1}{2}(\sigma^2_{h,t-1}\lambda_h + \varepsilon_t^2)h_t\right\} \ \mathrm{d}h_t, \ u := \frac{\sigma^2_{h,t-1}\lambda_h + \varepsilon_t^2}{2}$$

$$= \frac{2^{-\frac{1+\nu_h}{2}}(\sigma^2_{h,t-1}\lambda_h)^{\nu_h/2}}{\Gamma(\nu_h/2)\pi^{1/2}} \int h_t^{\frac{\nu_h-1}{2}} 2/(\sigma^2_{h,t-1}\lambda_h + \varepsilon_t^2)e^{-u} \ \mathrm{d}u$$

$$= \frac{2^{-\frac{1+\nu_h}{2}}(\sigma^2_{h,t-1}\lambda_h)^{\nu_h/2}}{\Gamma(\nu_h/2)\pi^{1/2}}\Gamma((\nu_h+1)/2)(\sigma^2_{h,t-1}\lambda_h + \varepsilon_t^2)^{-(\nu_h+1)/2}2^{(\nu_h+1)/2}$$

$$= \frac{\Gamma((\nu_h+1)/2)}{\Gamma(\nu_h/2)\pi^{1/2}}(\sigma^2_{h,t-1}\lambda_h)^{\nu_h/2}(\sigma^2_{h,t-1}\lambda_h + \varepsilon_t^2)^{-(\nu_h+1)/2}$$

$$= \frac{\Gamma((\nu_h+1)/2)}{\Gamma(\nu_h/2)\pi^{1/2}}(\sigma^2_{h,t-1}\lambda_h)^{-1/2}(1 + \varepsilon_t^2/(\sigma^2_{h,t-1}\lambda_h))^{-(\nu_h+1)/2},$$

which represents the density of a location-scale t-distribution with location $x'_t\beta_t$, scale $\sigma_{h,t-1}$ and ν_h degrees of freedom.

The volatility states $h_{1:T}$ which represent in this case the conditional precision matrices of the error term ($h_0 = 1/\sigma^2_{0,h}$ is fixed here) are sampled from their full conditional posterior given by

$$\pi\left(h_{1:T} \mid \theta_{-(\nu_h,\lambda_h)}, \gamma_{0:T}, y_{1:T}, \nu_h\right) \propto f_\theta(y_{1:T} \mid \gamma_{1:T}, h_{1:T}) \ f_\theta(h_{1:T}). \tag{B.1}$$

From Windle and Carvalho (2014, Propositions 1 and 2) it follows that the target density (B.1) can be easily simulated by backward sampling from the following Gamma distribution

$$h_t \mid (h_{t+1}, y_{1:T}) = \lambda_h h_{t+1} + Z_{t+1}, \qquad Z_{t+1} \sim \mathcal{G}(1/2, \sigma^2_{h,t+1}/2),$$

after forward filtering

$$\sigma^2_{t,h} = \lambda_h \sigma^2_{t-1,h} + \varepsilon_t^2.$$

6) Updating p_0:

Using inverse Gamma priors, i.e., for $i = 0, \ldots, n$,

$$p_{0,i} \sim \mathcal{IG}(\nu_p, (\nu_p - 1)c_p),$$

draws from the conditional posterior obtain as

$$\pi(p_{0,i} \mid \theta_{-(p_0)}, \gamma_{0:T}, y_{1:T}, h_{0:T}) \propto p(p_{0,i} \mid \gamma_{0,i}),$$

with $p_{0,i} \mid \gamma_{0,i} \sim \mathcal{IG}(\nu_p + \frac{1}{2}, (\nu_p - 1)c_p + \frac{1}{2}\gamma_{0,i}^2)$.

B.2.2. GMVP weight filtration for the TVP regression model

We are interested in an estimate $\hat{w}_t \mid \mathcal{F}_{t-1}, \hat{\theta}$ which we define as $\mathbb{E}(w_t \mid \mathcal{F}_{t-1}, \hat{\theta})$ setting the elements of $\hat{\theta}$ to their posterior mean. As $w_t = \beta_{t,1:n} + 1/n - 1/n \sum_{i=1}^{n} \beta_{t,1:n}$ and $\beta_t = \alpha + Q^{0.5}\gamma_t$, we need to find $f(\gamma_t \mid \mathcal{F}_{t-1}, \hat{\theta})$, and replace α and the elements of Q by their estimates. In case the error variance $\sigma_t^2 = \sigma^2$, with $\sigma^2 \in \theta$, $f(\gamma_t \mid \mathcal{F}_{t-1}, \hat{\theta})$ is normal with mean $\mu_{t|t-1}$ and covariance $V_{t|t-1}$ obtained by the prediction step of the Kalman filter. The Kalman filter for γ_t is initialized by $\mu_{1|0} = 0$ and $V_{1|0} = P_0 + I_{n+1}$. Then, denoting $Z_t = Q^{0.5}X_t$, the recursion is given as follows for $t = 1, \ldots, T$:

$$V_{t|t} = V_{t|t-1} - V_{t|t-1}Z_t(Z_t'V_{t|t-1}Z_t + \sigma_t^2)^{-1}Z_t'V_{t|t-1}$$
$$\mu_{t|t} = \mu_{t|t-1} + V_{t|t-1}Z_t(Z_t'V_{t|t-1}Z_t + \sigma_t^2)^{-1}(y_t - X_t'\alpha - Z_t'\mu_{t|t-1})$$
$$V_{t+1|t} = V_{t|t} + I_{n+1}$$
$$\mu_{t+1|t} = \mu_{t|t}.$$

In case σ_t^2 follows itself some stochastic process $\{h_t\}$, we need to integrate out the volatility states numerically. For that we make use of a marginalized bootstrap particle filter for $\{h_t\}$ using Rao-Blackwellization based on the Kalman filter for $\{\gamma_t\}$. The following algorithm is based on Schön et al. (2005). Conditioning on the parameters is omitted for notational convenience:

Algorithm for Rao-Blackwellized particle filter

$t = 1$

- for $i = 1, \ldots, S$

    - sample $h_1^{(i)} \sim f(h_1)$.

    - obtain KF update $V_{1|1}^{(i)}$ and $\mu_{1|1}^{(i)}$ based on the respective draws for $h_1^{(i)}$ ($\sigma_t^2 = \exp(h_t)$).

    - compute

    $$p(y_1 \mid h_1^{(i)}) = \frac{p(y_1 \mid \gamma_1, h_1^{(i)}) \, p(\gamma_1 \mid h_1^{(i)})}{p(\gamma_1 \mid h_1^{(i)}, y_1)}$$

with

$$y_1 \mid \gamma_1, h_1^{(i)} \sim \mathcal{N}(X_1'\alpha + Z_1'\gamma, \sigma_1^{2(i)}),$$
$$\gamma_1 \mid h_1^{(i)} \sim \mathcal{N}(\mu_{1|0}, V_{1|0}), \quad \gamma_1 \mid h_1^{(i)}, y_1 \sim \mathcal{N}(\mu_{1|1}^{(i)}, V_{1|1}^{(i)}),$$

which leads to

$$p(y_1 \mid h_1^{(i)}) = \frac{1}{\sqrt{2\pi\sigma_1^{2(i)}}} \left(\frac{|V_{1|1}^{(i)}|}{|V_{1|0}|} \right)^{1/2}$$
$$\times \exp\left\{ -\frac{1}{2} \left(\frac{(y_1 - X_1'\alpha)^2}{\sigma_1^{2(i)}} + \mu_{1|0}'V_{1|0}^{-1}\mu_{1|0} - \mu_{1|1}^{(i)}{}'V_{1|1}^{(i)-1}\mu_{1|1}^{(i)} \right) \right\}.$$

Rearranging reveals that this implies

$$y_1 \mid h_1^{(i)} \sim \mathcal{N}(\mu_1^{*(i)}, \sigma_1^{2*(i)}),$$

with $\mu_1^{*(i)} = Z_1'\mu_{1|0} + X_1'\alpha$, and $\sigma_1^{2*(i)} = Z_1'V_{1|0}Z_1 + \sigma_1^{2(i)}$.

- obtain non-normalized weights $w_1^{(i)} = p(y_1 \mid h_1^{(i)})$ and normalized weights $\tilde{w}_1^{(i)} = \frac{w_1^{(i)}}{\sum_{i=1}^{S} w_1^{(i)}}$.

- for $i = 1, \ldots, S$: resample $j_i \sim \text{Multinomial}(\tilde{w}_1^{(1)}, \ldots, \tilde{w}_1^{(S)})$ and set

$$\left\{ h_1^{(i)} = h_1^{(j_i)}, \quad \mu_{1|1}^{(i)} = \mu_{1|1}^{(j_i)}, \quad V_{1|1}^{(i)} = V_{1|1}^{(j_i)} \right\}_{i=1}^{S}.$$

$t = 2, \ldots, T$

- for $i = 1, \ldots, S$

  – sample $h_t^{(i)} \sim f(h_t \mid h_{t-1}^{(i)})$ and obtain $h_{1:t}^{(i)} = (h_{1:t-1}^{(i)}, h_t^{(i)})$.

  – obtain KF prediction $V_{t|t-1}^{(i)}$ and $\mu_{t|t-1}^{(i)}$.

  – obtain KF update $V_{t|t}^{(i)}$ and $\mu_{t|t}^{(i)}$.

  – compute

$$p(y_t \mid h_{1:t}^{(i)}, y_{1:t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^{2(i)}}} \left(\frac{|V_{t|t}^{(i)}|}{|V_{t|t-1}^{(i)}|} \right)^{1/2}$$
$$\times \exp\left\{ -\frac{1}{2} \left(\frac{(y_t - X_t'\alpha)^2}{\sigma_t^{2(i)}} + \mu_{t|t-1}^{(i)}{}'V_{t|t-1}^{(i)}{}^{-1}\mu_{t|t-1}^{(i)}{}' - \mu_{t|t}^{(i)}V_{t|t}^{(i)-1}\mu_{t|t}^{(i)} \right) \right\}$$

which implies

$$y_t \mid h_{1:t}^{(i)}, y_{1:t-1} \sim \mathcal{N}(\mu_t^{*(i)}, \sigma_t^{2*(i)}),$$

with $\mu_t^{*(i)} = Z_t' \mu_{t|t-1}^{(i)} + X_t' \alpha$ and $\sigma_t^{2*(i)} = Z_t' V_{t|t-1}^{(i)} Z_t + \sigma_t^{2(i)}$.

- obtain non-normalized weights $w_t^{(i)} = p(y_t \mid h_{1:t}^{(i)}, y_{1:t-1})$ and normalized weights $\tilde{w}_t^{(i)} = \dfrac{w_t^{(i)}}{\sum_{i=1}^{S} w_t^{(i)}}$.

- for $i = 1, \ldots, S$: resample $j_i \sim \text{Multinomial}(\tilde{w}_t^{(1)}, \ldots, \tilde{w}_t^{(S)})$ and set

$$\left\{ h_t^{(i)} = h_t^{(j_i)}, \quad \mu_{t|t}^{(i)} = \mu_{t|t}^{(j_i)}, \quad V_{t|t}^{(i)} = V_{t|t}^{(j_i)} \right\}_{i=1}^{S}.$$

After re-sampling, $f(\gamma_t \mid y_{1:t})$ and $f(\gamma_{t+1} \mid y_{1:t})$ can be approximated for each point in time as

$$\hat{f}(\gamma_t \mid y_{1:t}) = \frac{1}{S} \sum_{i=1}^{S} f(\gamma_t \mid y_{1:t}, h_{1:t}^{(i)}), \qquad \hat{f}(\gamma_{t+1} \mid y_{1:t}) = \frac{1}{S} \sum_{i=1}^{S} f(\gamma_{t+1} \mid y_{1:t}, h_{1:t}^{(i)}),$$

which leads to the following approximation:

$$\widehat{\mathbb{E}(\gamma_{t+1} \mid \mathcal{F}_t, \hat{\theta})} = \frac{1}{S} \sum_{i=1}^{S} \mu_{t+1|t}^{(i)} = \frac{1}{S} \sum_{i=1}^{S} \mu_{t|t}^{(i)}.$$

B.3. Alternative volatility specifications

In this section, alternative volatility specifications and corresponding out-of-sample results are presented. Tables regarding these alternative specifications are marked by the note '(alt)' in the captions.

B.3.1. Taylor (1982) model with random walk for $\log(\sigma_t^2)$

As alternative approach to incorporate stochastic volatility in our model, we impose for the transition of the log squared volatilities a random walk, i.e., the persistence parameter ϕ_h in (3.10) fixed to one and different initial conditions:

$$h_t = h_{t-1} + \eta_t^h, \qquad \eta_t^h \overset{\text{iid}}{\sim} \mathcal{N}(0, R), \qquad h_0 \sim \mathcal{N}(\mu_0^h, V_0^h).$$

Particularly, we set $\mu_0^h = 0$, $V_0^h = 1$.

B.3.2. Beta process for $1/\sigma_t^2$ (Uhlig, 1994, 1997)

As alternative approach to incorporate stochastic volatility in our model, we employ a univariate version of the specification of Windle and Carvalho (2014) which generalizes the approach of Uhlig (1994) and Uhlig (1997). It specifies the transition of the precision of the errors, i.e., $\sigma_t^2 = 1/h_t$, imposing the following Beta process:

$$h_t = \frac{1}{\lambda_h} h_{t-1} \psi_t, \qquad\qquad \psi_t \sim \mathcal{B}_1\left(\frac{n_h}{2}, \frac{1}{2}\right),$$

with parameters $n_h, \lambda_h > 0$ and initial condition

$$h_1 \sim \mathcal{G}\left(n_h/2, \lambda_h \sigma_{0,h}^2/2\right).$$

The parameter n_h determines the conditional variation of $h_t \mid h_{t-1}$, i.e., a large value of n_h leads to a smaller variation in the innovation ψ_t for a given level λ_h. We use a flat prior for λ_h, $\lambda_h \propto 1/\lambda_h$ and fix $\sigma_h^2 = 1$. We sample λ_h, n_h (jointly) by a standard Gaussian RW-MH-step.

B.3.3. Empirical results

In Tables B.1 and B.2, we present out-of-sample forecasting results for the previously discussed alternative volatility specifications compared to the stationary AR(1) SV-model and the specification without stochastic volatility, that are employed in the main version of the paper. For the experiments we use the same data set and design, as well as asset universe selection as employed in the main paper, under the restriction that the sample covers only the time span until 2016. Precisely, the sample covers the period from 01/07/2002 to 12/23/2016 for a total of T = 3770 trading days. The out-of-sample period spans the period 12/21/2006 to 12/23/2016 for a total of $21 \times 120 = 2520$ trading days. We denote the SV specification outlined in Section B.3.1 as *TA-RW* and the specification of Section B.3.2 as *UH*.

Table B.1.: Out-of-sample GMVP variance (alt.)

	$n = 100$	$n = 200$	$n = 400$
$T = 250$			
TVP-GMVP-			
TA-RW-SV-*shr*	0.617	0.525	**0.462**
TA-RW-SV-*no shr*	1.970	27.904	14.040
UH-SV-*shr*	**0.593**	**0.523**	0.511
UH-SV-*no shr*	2.301	22.120	3.810
SV-*shr*	0.648	0.577	0.487
SV-*no shr*	0.864	2.044	4.063
no SV-*shr*	0.624	0.538	0.484
no SV-*no shr*	2.402	23.204	1.499
$T = 1250$			
TVP-GMVP-			
TA-RW-SV-*shr*	0.611	0.586	**0.476**
TA-RW-SV-*no shr*	0.580	0.667	0.588
UH-SV-*shr*	0.580	**0.527**	0.518
UH-SV-*no shr*	**0.566**	0.603	0.594
SV-*shr*	0.608	0.570	0.484
SV-*no shr*	0.631	0.607	0.560
no SV-*shr*	0.627	0.596	0.569
no SV-*no shr*	0.593	0.955	0.634

Note: Out-of-sample model comparison. Smallest value in bold letters. Grey light (dark) shaded cells indicate that the model belongs to the 90 (75)% MCS.

Table B.2.: Out-of-sample Sharpe ratio ($\times$10) (alt.)

	$c = 0$ bp.			$c = 10$ bp.		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-						
TA-RW-SV-*shr*	0.606	0.509	0.390	0.532	0.429	0.312
TA-RW-SV-*no shr*	0.468	0.021	0.148	0.311	-0.284	-0.222
UH-SV-*shr*	0.538	0.466	0.345	0.486	0.408	0.221
UH-SV-*no shr*	0.433	0.085	0.331	0.186	-0.283	0.101
SV-*shr*	0.615	**0.533**	0.358	0.524	0.429	0.253
SV-*no shr*	**0.639**	0.324	0.179	0.492	0.026	-0.281
no SV-*shr*	0.629	0.518	**0.406**	**0.548**	**0.432**	**0.324**
no SV-*no shr*	0.282	0.184	0.068	0.071	-0.128	-0.088

T = 250 is the header spanning the first block above.

	$c = 0$ bp.			$c = 10$ bp.		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
TVP-GMVP-						
TA-RW-SV-*shr*	0.518	0.391	0.379	0.462	0.324	**0.312**
TA-RW-SV-*no shr*	0.535	0.412	0.299	0.489	0.229	0.080
UH-SV-*shr*	0.535	**0.453**	**0.406**	0.489	**0.409**	0.292
UH-SV-*no shr*	0.526	0.342	0.221	0.491	0.223	0.082
SV-*shr*	0.530	0.398	0.395	0.475	0.325	0.303
SV-*no shr*	0.560	0.389	0.275	0.490	0.297	0.151
no SV-*shr*	**0.591**	0.392	0.307	0.517	0.323	0.188
no SV-*no shr*	0.582	0.183	0.221	**0.531**	0.028	0.076

T = 1250 is the header spanning the second block above.

Note: GMVP weights for TVP regression specifications. Largest value in bold letters. Results are displayed for $c = 0$ and $c = 10$ basis points trading costs.

Appendix C.

Appendix for Chapter 4

C.1. Precision matrix and conditional correlation

For showing that conditional correlations in a Gaussian model can be expressed by the elements of the precision, I exploit the rules for block-wise matrix inversion and partition, without loss of generality, the covariance matrix Σ of a set of variables $X = (X_1, \ldots, X_n)'$ with $X \sim N(\mu, \Sigma)$ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with Σ_{11} denoting $\mathbb{V}[X_1, \ldots, X_k]$ and $\Sigma_{22} = \mathbb{V}[X_{k+1}, \ldots, X_n]$. Then I can express the analogously block-partitioned precision matrix $\Theta = \Sigma^{-1}$ as

$$\Theta = \begin{pmatrix} A^{-1} & -A^{-1}C \\ -BA^{-1} & \Sigma_{22}^{-1} + BA^{-1}C \end{pmatrix},$$

$$A = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \quad B = \Sigma_{22}^{-1}\Sigma_{21}, \quad C = \Sigma_{12}\Sigma_{22}^{-1}.$$

Next, I denote that the conditional covariance $\mathbb{V}[X_1, \ldots, X_k \mid X_{k+1}, \ldots X_n]$ under a multivariate normal distribution is given as $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ which corresponds exactly to A, the inverse of the $k \times k$ block in Θ corresponding to entries $1, \ldots, k$. Setting, e.g., $k = 2$, the conditional correlation of X_1, X_2 given $X_3, \ldots, X_n$ is hence given by

$$\rho_{1,2|3,\ldots,n} = \frac{A_{[12]}}{\sqrt{A_{[11]}A_{[22]}}},$$

which simplifies, by calculating the inverse of the 2×2 matrix A, to

$$\rho_{1,2|3,\ldots,n} = \frac{-\Theta_{[12]}}{\sqrt{\Theta_{[11]}\Theta_{[22]}}},$$

which holds true analogously for any other $i, j \in (1, \ldots, n), i \neq j$. Accordingly, iff some element $\Theta_{[ij]}$ is equal to zero, it follows that X_i is conditionally independent of X_j (see also

Lauritzen, 1996, Hastie et al., 2009).

C.2. Proximal operator

Proximal operators (Parikh and Boyd, 2014) define a class of particular optimization problems of the form

$$\text{prox}_{\eta, f(X)}(A) = \underset{X \in R^{m \times n}}{\text{argmin}} \left(f(X) + 1/(2\eta) ||X - A||_2^2 \right),$$

for some real matrix $A \in R^{m \times n}$ and a real-valued function $f(X)$, which is proper, convex and at least semi-continuous. They define a trade-off for X between minimizing f and being close to A. In the ADMM algorithm applied for the ATVGL proposed in this paper, two particular proximal operators with well-known closed form solutions are being used:

1.

$$\text{prox}_{\eta, -\log(\det(X)) - \text{trace}(SX)}((A + A')/2) = \frac{\eta}{2} Q(D + \sqrt{D^2 + 4\eta^{-1}I_m} Q', \qquad \text{(C.1)}$$

   with diagonal D and orthonormal Q representing the components of the eigenvalue decomposition $\frac{A + A'}{2\eta} - S = QDQ'$ (Witten et al., 2009).

2.

$$\text{prox}_{\eta, ||X||_1}(A) = Y, \quad \text{with } Y_{[ij]} = \begin{cases} 0, & \text{if } |A_{[ij]}| \leq \eta \\ \text{sign}(A_{[ij]})(|A_{[ij]}| - \eta) & \text{else,} \end{cases} \qquad \text{(C.2)}$$

   which is known as the element-wise soft thresholding operator (Parikh and Boyd, 2014, Ch.6.5.2).

C.3. Optimization steps

C.3.1. Minimization of $\mathcal{L}_\rho$ with respect to Θ_τ

Analogously to Boyd et al. (2011), I derive the solution for minimizing the negative of the penalized log-likelihood $\mathcal{L}_\rho$ with respect to $\bar{\Theta}$ individually for each Θ_τ:

$$\underset{\Theta_\tau \in S_n^{++}}{\text{argmin}} \mathcal{L}_\rho(\Theta, Z, U) = \underset{\Theta_\tau \in S_n^{++}}{\text{argmin}} -\log(\det(\Theta_\tau)) - \text{trace}(S_\tau \Theta_\tau) + \frac{\rho k}{2\mathcal{N}_\tau} ||(\Theta_\tau - A^\tau))||_2^2,$$

$$\text{(C.3)}$$

where $A^\tau = \frac{\sum_{j=1}^{J} Z_{\tau,0,j}^k + \sum_{m=1}^{M}(Z_{\tau,1,m}+Z_{\tau,2,m}) - \sum_{j=1}^{J} U_{\tau,0,j}^k - \sum_{m=1}^{M}(U_{\tau,1,m}+U_{\tau,2,m})}{k}$ and $k = 2M + J$. Equation (C.3) has the form of a proximal operator with

$$f(\Theta_\tau) := -\log(\det(\Theta_\tau)) - \text{trace}(S_\tau \Theta_\tau),$$
$$\eta := \frac{\mathcal{N}_\tau}{\rho k},$$
$$A := A^\tau.$$

Rewriting the symmetric matrix $A^\tau = \frac{A^\tau + A^{\tau\prime}}{2}$ with eigendecomposition $\frac{1}{\frac{\mathcal{N}_\tau}{\rho k}}(A^\tau + A^{\tau\prime})/2 - S_\tau = Q^\tau D^\tau Q^{\tau\prime}$, leads, according to Equation (C.1), to

$$\underset{\Theta_\tau \in S_n^{++}}{\text{argmin}}\, \mathcal{L}_\rho(\Theta, Z, U) = \text{prox}_{\frac{\mathcal{N}_\tau}{\rho k}, -\log(\det(\Theta_\tau))+\text{trace}(S_\tau \Theta_\tau)}(A^\tau)$$
$$= \frac{\mathcal{N}_\tau}{\rho k} Q^\tau \left(D^\tau + \sqrt{D^{\tau 2} + 4\frac{1}{2\frac{\mathcal{N}_\tau}{\rho k}} I_n} \right) Q^{\tau\prime}.$$

C.3.2. Minimization of $\mathcal{L}_\rho$ with respect to $Z_{\tau,0,1}$ and $(Z_{\tau-1,1,1}, Z_{\tau,2,1})$ $(\Psi_1, \tilde{\Psi}_1)$

In the following, I sketch the derivation of the solution for minimizing the augmented Lagrangian $\mathcal{L}_\rho$ with respect to $Z_{\tau,0,1}$ as well as $(Z_{\tau-1,1,1}, Z_{\tau,2,1})$ as derived in Boyd et al. (2011). The $Z_{\tau,0,1}$ update is given as

$$\underset{(Z_{\tau,0,1})}{\text{argmin}}\, \mathcal{L}_\rho(\bar{\Theta}, Z, U) = \underset{(Z_{\tau,0,1})}{\text{argmin}}\, \lambda_1 \sum_{i \neq j} \sum_j |(Z_{\tau,0,1,[ij]})| + ||\,(Z_{\tau,0,1} - (\Theta_\tau + U_{\tau,0,1}))\,||_2^2,$$

which directly defines a proximal operator analogously to Equation (C.2) with

$$f(Z_{\tau,0,1}) := ||(Z_{\tau,0,1})||_{1,odd}$$
$$\eta := \frac{\lambda_1}{\rho},$$
$$A := (\Theta_\tau + U_{\tau,0,1}),$$

where $||\cdot||_{1,odd}$ denotes the element-wise L_1 norm for a matrix omitting its diagonals. Hence, the solution is given by the element-wise soft thresholding operator as:

$$\underset{(Z_{\tau,0,1})}{\text{argmin}}\, \mathcal{L}_\rho(\bar{\Theta}, Z, U) = \text{prox}_{\frac{\lambda_1}{\rho}, ||(Z_{\tau,0,1})||_{1,odd}}(\Theta_\tau + U_{\tau,0,1}) = Y^\tau, \quad \text{with}$$

$$Y_{[ij]}^\tau = \begin{cases} (\Theta_\tau + U_{\tau,0,1})_{[ij]} & \text{if } i = j \\ 0 & \text{if } i \neq j \wedge |Z_{\tau,0,1,[ij]}| \leq \frac{\lambda_1}{\rho} \\ \text{sign}((\Theta_\tau + U_{\tau,0,1})_{[ij]})(|(\Theta_\tau + U_{\tau,0,1})_{[ij]}| - \eta) & \text{if } i \neq j \wedge |(\Theta_\tau + U_{\tau,0,1})_{[ij]}| > \frac{\lambda_1}{\rho}. \end{cases}$$

The $(Z_{\tau-1,1,1}, Z_{\tau,2,1})$ update is given as

$$\underset{\text{vec}(Z_{\tau-1,1,m},Z_{\tau,2,m})}{\operatorname{argmin}} \mathcal{L}_\rho(\bar{\Theta}, Z, U)$$
$$= \underset{\text{vec}(Z_{\tau-1,1,m},Z_{\tau,2,m})}{\operatorname{argmin}} \beta_1 \sum_{i,j} |(Z_{\tau-1,1,1,[ij]} - Z_{\tau,2,1,[ij]})|$$
$$+ \rho/2(\| (Z_{\tau-1,1,1} - (\Theta_{\tau-1} + U_{\tau-1,1,1})) \|_2^2 - \| (Z_{\tau,2,1} - (\Theta_\tau + U_{\tau,2,1})) \|_2^2),$$

which defines a proximal operator of the form

$$\operatorname{prox}_{\frac{\beta_1}{\rho}, \|(Z_{\tau-1,1,m}-Z_{\tau,2,m})\|_1} \left(\begin{bmatrix} \Theta_{\tau-1} + U_{\tau-1,1,1} \\ \Theta_\tau + U_{\tau,2,1} \end{bmatrix} \right).$$

Exploiting functional properties of the element-wise norm, this can be converted into

$$\begin{bmatrix} Z_{\tau-1,1,1} \\ Z_{\tau,2,1} \end{bmatrix} = 1/2 \begin{bmatrix} \Theta_{\tau-1} + U_{\tau-1,1,1} - \Theta_\tau + U_{\tau,2,1} \\ \Theta_{\tau-1} + U_{\tau-1,1,1} - \Theta_\tau + U_{\tau,2,1} \end{bmatrix} + 1/2 \begin{bmatrix} -E^\tau \\ +E^\tau \end{bmatrix},$$

with E^τ given as solution of the element-wise soft thresholding proximal operator with

$$f((Z_{\tau-1,1,m}, Z_{\tau,2,m})) := \|((Z_{\tau-1,1,m} - Z_{\tau,2,m}))\|_1$$
$$\eta := \frac{2\beta_1}{\rho},$$
$$A := \Theta_{\tau-1} + U_{\tau-1,1,1} - \Theta_\tau + U_{\tau,2,1}.$$

C.3.3. Minimization of $\mathcal{L}_\rho$ with respect to $Z_{\tau,0,2}$ (Ψ_2)

I now illustrate the solution for minimizing the augmented Lagrangian $\mathcal{L}_\rho$ with respect to $Z_{\tau,0,2}$. First I denote that

$$\Psi_2(X) = \|(X\iota - \iota'X\iota/n)\|_2^2,$$

after rearranging can be written as

$$\Psi_2(X) = \operatorname{trace}(X1_{n\times n}X) - \operatorname{trace}(\text{vec}(X)'1_{n^2\times n^2}\text{vec}(X)/n).$$

This implies the following optimization problem of the augmented Lagrangian:

$$\underset{(Z_{\tau,0,2})}{\operatorname{argmin}} \mathcal{L}_\rho(\bar{\Theta}, Z_{\tau,0,2}, U) = \underset{(Z_{\tau,0,2})}{\operatorname{argmin}} \lambda_2 \left(\operatorname{trace}(Z_{\tau,0,2}1_{n\times n}Z_{\tau,0,2})\right.$$
$$-\operatorname{trace}(\text{vec}(Z_{\tau,0,2})'1_{n^2\times n^2}\text{vec}(Z_{\tau,0,2})/n)$$
$$+ \rho/2\| (Z_{\tau,0,2} - (\Theta_\tau + U_{\tau,0,2})) \|_2^2.$$

Taking first order conditions with respect to $\text{vec}(Z_{\tau,0,2})$ leads to

$$\frac{\partial \mathcal{L}_\rho(\bar{\Theta}, Z_{\tau,0,2}, U)}{\partial \text{vec}(Z_{\tau,0,2})}$$

$$= \left(\underbrace{\lambda_2[(1_{n \times n} \otimes I_n) - (I_n \otimes 1_{n \times n}) - 2/n 1_{n^2 \times n^2}] + \rho I_{n^2}}_{:=J_{\Psi_2}} \right) \text{vec}(Z_{\tau,0,2}) \overset{!}{=} \rho \text{vec}(\Theta_\tau + U_{\tau,0,2}),$$

and hence the solution $\text{vec}(Z^*_{\tau,0,2})$ is given by

$$\text{vec}(Z^*_{\tau,0,2}) = (J_{\Psi_2})^{-1} \rho \text{vec}(\Theta_\tau + U_{\tau,0,2}).$$

Notably, J_{Ψ_2} depends only on n, ρ and λ_2 such that its inversion only needs to be performed once for a parameter setting before starting the ADMM. However, the $n^2 \times n^2$ matrix $J_{\Psi_2}^{-1}$ is not sparse such that the multiplication with the $n^2 \times 1$ vector $\rho \text{vec}(\Theta_\tau + U_{\tau,0,2})$ which has to be performed for all τ and all iterations k, can become very slow for large dimensions. Moreover, the required memory for J_{Ψ_2} and its inverse grows at rate $\mathcal{O}(n^4)$; for $n = 100$ it is equal to 1.6GB, and for $n = 200$, already 25.6GB are required. Hence, I exploit the structure of J_{Ψ_2} to derive an analytical expression for its inverse:

$$J_{\Psi_2} = \begin{pmatrix} A_1 & A_2 & A_2 & \dots & A_2 \\ A_2 & A_1 & A_2 & \dots & A_2 \\ & & \ddots & & \\ & & & \ddots & \\ A_2 & \dots & & A_2 & A_1 \end{pmatrix},$$

where the $n \times n$ blocks A_1 and A_2 are given by

$$A_1 = \begin{pmatrix} a_1 & a_2 & a_2 & \dots & a_2 \\ a_2 & a_1 & a_2 & \dots & a_2 \\ & & \ddots & & \\ & & & \ddots & \\ a_2 & \dots & & a_2 & a_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & a_3 & a_3 & \dots & a_3 \\ a_3 & a_2 & a_3 & \dots & a_3 \\ & & \ddots & & \\ & & & \ddots & \\ a_3 & \dots & & a_3 & a_2 \end{pmatrix},$$

with

$$a_1 = (2 - 2/n)\lambda_2 + \rho,$$
$$a_2 = (1 - 2/n)\lambda_2,$$
$$a_3 = -2/n\lambda_2.$$

Hence J_{Ψ_2} consists of three distinct entries only. Likewise, $J_{\Psi_2}^{-1}$ has only three different elements. It is given by

$$
J_{\Psi_2}^{-1} = \begin{pmatrix} B_1 & B_2 & B_2 & \dots & B_2 \\ B_2 & B_1 & B_2 & \dots & B_2 \\ & & \ddots & & \\ & & & \ddots & \\ B_2 & \dots & & B_2 & B_1 \end{pmatrix},
$$

where the $n \times n$ blocks B_1 and B_2 are given by

$$
B_1 = \begin{pmatrix} b_1 & b_2 & b_2 & \dots & b_2 \\ b_2 & b_1 & b_2 & \dots & b_2 \\ & & \ddots & & \\ & & & \ddots & \\ b_2 & \dots & & b_2 & b_1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} b_2 & b_3 & b_3 & \dots & b_3 \\ b_3 & b_2 & b_3 & \dots & b_3 \\ & & \ddots & & \\ & & & \ddots & \\ b_3 & \dots & & b_3 & b_2 \end{pmatrix},
$$

with b_1 b_2 and b_3 being the following functionals of ρ, λ_2 and n:

$$
b_1 = 1/n((n^2 - 2n + 2)\lambda_2 + n\rho)/(\rho(n\lambda_2 + \rho)),
$$
$$
b_2 = -(n-2)\lambda_2/(n\rho(n\lambda_2 + \rho)),
$$
$$
b_3 = (2\lambda_2)/(n\rho(n\lambda_2 + \rho)).
$$

Using this result, I can circumvent the vast-dimensional matrix multiplication and replace it by computing appropriately weighted sums of the elements of $\rho\mathrm{vec}(\Theta_\tau + U_{\tau,0,2})$, which reduces computing time by more than 98% and makes the penalty function Ψ_2 scalable to large and vast dimensional applications.

C.4. Additional empirical results

Table C.1.: Out-of-sample MVP Sharpe ratio ($\times 10$) for $n = 100$

model	unrestricted	130/30	100/0
ATVGL	**0.484**	0.415	0.307
DCC-nl	0.435	0.380	0.280
DCC	0.433	0.376	0.282
AFM-DCC-nl	0.439	0.394	0.292
SHR-nl	0.423	0.363	0.287
SHR-nl*	0.417	0.330	0.285
SHR-l	0.423	0.337	0.282
SHR-l-1f	0.427	0.335	0.281
plug-in	0.423	0.335	0.282
naïve	0.444	**0.444**	**0.444**

Note: Out-of-sample model comparison MV-with momentum signal portfolio. Largest value in bold letters.

Table C.2.: Out-of-sample empirical variance unrestricted GMVP for $n = 200$

model/time	full	'80-'84	'85-'89	'90-'94	'95-'99	'00-'04	'05-'09	'10-'14	'15-'19
ATVGL	0.393	0.272	0.535	0.196	0.358	0.577	0.567	0.293	0.347
DCC-nl	0.410	0.274	0.515	0.234	0.368	0.635	0.560	0.310	0.383
DCC	0.437	0.296	0.568	0.254	0.395	0.668	0.584	0.328	0.402
AFM-DCC-nl	0.424	0.275	0.553	0.247	0.377	0.641	0.581	0.318	0.395
SHR-nl	0.442	0.283	0.606	0.248	0.392	0.674	0.588	0.343	0.396
SHR-nl*	0.418	0.289	0.574	0.213	0.394	0.614	0.577	0.311	0.373
SHR-l	0.466	0.305	0.654	0.268	0.414	0.690	0.613	0.357	0.421
SHR-l-1f	0.444	0.284	0.599	0.257	0.395	0.679	0.590	0.339	0.402
plug-in	0.467	0.305	0.656	0.270	0.415	0.691	0.616	0.361	0.422
RLS-REF	0.424	0.293	0.587	0.229	0.387	0.611	0.599	0.314	0.365
naïve	1.157	0.826	1.312	0.570	0.925	1.541	2.321	1.055	0.699

Note: Grey shaded cells indicate that the model belongs to the 90 % MCS.

Table C.3.: Out-of-sample empirical variance exposure-constrained GMVP for $n = 200$

model/time	full	'80-'84	'85-'89	'90-'94	'95-'99	'00-'04	'05-'09	'10-'14	'15-'19
				130/30 strategy					
ATVGL	0.402	0.288	0.563	0.199	0.365	0.582	0.560	0.296	0.363
DCC-nl	0.407	0.288	0.538	0.225	0.375	0.603	0.549	0.316	0.358
DCC	0.412	0.290	0.556	0.228	0.380	0.608	0.554	0.316	0.357
AFM-DCC-nl	0.442	0.281	0.588	0.251	0.386	0.635	0.630	0.349	0.408
SHR-nl	0.457	0.286	0.621	0.264	0.401	0.694	0.627	0.350	0.411
SHR-nl*	0.416	0.290	0.607	0.207	0.386	0.590	0.572	0.305	0.364
SHR-l	0.461	0.290	0.632	0.270	0.404	0.691	0.634	0.349	0.412
SHR-l-1f	0.457	0.287	0.607	0.276	0.400	0.691	0.629	0.346	0.415
plug-in	0.461	0.291	0.632	0.271	0.404	0.691	0.634	0.349	0.412
naïve	1.157	0.826	1.312	0.570	0.925	1.541	2.321	1.055	0.699
				100/0 strategy					
ATVGL	0.487	0.363	0.667	0.245	0.420	0.656	0.771	0.379	0.394
DCC-nl	0.459	0.369	0.601	0.246	0.430	0.651	0.636	0.373	0.360
DCC	0.460	0.367	0.609	0.247	0.435	0.652	0.638	0.373	0.359
AFM-DCC-nl	0.518	0.373	0.659	0.281	0.436	0.695	0.828	0.428	0.440
SHR-nl	0.536	0.384	0.690	0.301	0.450	0.766	0.831	0.434	0.431
SHR-nl*	0.490	0.370	0.679	0.248	0.431	0.651	0.766	0.389	0.385
SHR-l	0.539	0.384	0.699	0.309	0.453	0.760	0.837	0.436	0.434
SHR-l-1f	0.535	0.384	0.673	0.311	0.448	0.763	0.833	0.432	0.434
plug-in	0.539	0.384	0.698	0.309	0.453	0.760	0.838	0.436	0.434
naïve	1.157	0.826	1.312	0.570	0.925	1.541	2.321	1.055	0.699

Note: Grey shaded cells indicate that the model belongs to the 90 % MCS.

Table C.4.: Out-of-sample MVP Sharpe ratio ($\times 10$) for $n = 200$

model	unrestricted	130/30	100/0
ATVGL	**0.644**	**0.487**	0.370
DCC-nl	0.448	0.424	0.360
DCC	0.447	0.423	0.359
AFM-DCC-nl	0.457	0.427	0.378
SHR-nl	0.441	0.421	0.353
SHR-nl*	0.424	0.333	0.313
SHR-l	0.443	0.350	0.311
SHR-l-1f	0.448	0.349	0.311
plug-in	0.443	0.349	0.311
naïve	0.478	0.478	**0.478**

Note: Out-of-sample model comparison MV-with momentum signal portfolio. Largest value in bold letters.

Bibliography

Ait-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Amemiya, T. (1985). *Advanced Econometrics*. Havard University Press.

Angelosante, D., Bazerque, J. A., and Giannakis, G. B. (2010). Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1 -norm. *IEEE Transactions on Signal Processing*, 58:3436–3447.

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

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

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

Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. OUP Oxford.

Belmonte, M. A., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94.

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

Bittanti, S., Bolzern, P., and Campi, M. (1990). Convergence and exponential convergence of identification algorithms with directional forgetting factor. *Automatica*, 26:929–932.

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97.

Bodnar, T., Parolya, N., and Thorsen, E. (2021). Dynamic shrinkage estimation of the high-dimensional minimum-variance portfolio. *arXiv preprint arXiv:2106.02131*.

Boyd, S., Parikh, N., and Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Brandt, M. W. (2010). Portfolio choice problems. In Ait-Sahalia, Y. and Hansen, L., editors, *Handbook of Financial Econometrics*, pages 269–336. Elsevier.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

Bibliography

Callot, L., Caner, M., Önder, A. Ö., and Ulaşan, E. (2019). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32:140–158.

Candelon, B., Hurlin, C., and Tokpavi, S. (2012). Sampling error and double shrinkage estimation of minimum variance portfolios. *Journal of Empirical Finance*, 19:511–527.

Caner, M., Medeiros, M., and Vasconcelos, G. (2020). Sharpe ratio in high dimensions: Cases of maximum out of sample, constrained maximum, and optimal portfolio choice. *arXiv preprint arXiv:2002.01800*.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Clements, A., Scott, A., and Silvennoinen, A. (2015). On the benefits of equicorrelation for portfolio allocation. *Journal of Forecasting*, 34:507–522.

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

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28:777–795.

Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens conference on applied probability and time series analysis*, pages 145–159. Springer.

De Nard, G. (2020). Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage. *Journal of Financial Econometrics*.

De Nard, G., Ledoit, O., and Wolf, M. (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., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 5:798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22:1915–1953.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263.

Elliott, G., Ghanem, D., and Krüger, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Review of Economics and Statistics*, 98:742–755.

146

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

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30:212–228.

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

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

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4).

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107:592–606.

Ferson, W. E. and Siegel, A. F. (2001). The efficient use of conditioning information in portfolios. *The Journal of Finance*, 56:967–982.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44:1680–1707.

Fissler, T. and Ziegel, J. F. (2019). Order sensitivity and equivariance of scoring functions. *Electronic Journal of Statistics*, 13:1166–1211.

French, K. R. (2008). Presidential address: The cost of active investing. *The Journal of Finance*, 63(4):1537–1573.

Frey, C. and Pohlmeier, W. (2016). Bayesian shrinkage of portfolio weights. Working paper, University of Konstanz.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25:595–620.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746–762.

Goel, A. and Bernstein, D. S. (2018). A targeted forgetting factor for recursive least squares. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3899–3903.

Golosnoy, V., Gribisch, B., and Seifert, M. I. (2019). Exponential smoothing of realized portfolio weights. *Journal of Empirical Finance*, 53:222–237.

Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1.

Hallac, D., Park, Y., Boyd, S., and Leskovec, J. (2017). Network inference via the time-varying graphical lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 205–213.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23:365–380.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79:453–497.

Harvey, A. C. (1993). *Time Series Models*. Pearson.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hörmann, W. and Leydold, J. (2014). Generating generalized inverse gaussian random variates. *Statistics and Computing*, 24(4):547–557.

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

Janková, J. and van de Geer, S. (2018). Inference in high-dimensional graphical models. *Handbook of Graphical Models*, page 325.

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

Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

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

Kempf, A. and Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review*, 58:332–348.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

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

Kulhavỳ, R. and Zarrop, M. B. (1993). On a general concept of forgetting. *International Journal of Control*, 58:905–924.

Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. ACM.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

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

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

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

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

Ledoit, O. and Wolf, M. (2020). The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*.

Lee, T.-H. and Seregina, E. (2021). Optimal portfolio using factor graphical lasso. *arXiv preprint arXiv:2011.00435*.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, 33(3):381–392.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, pages 13–37.

Ljung, L. and Söderström, T. (1983). *Theory and Practice of Recursive Identification*. MIT Press.

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

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state–space models: A computational efficiency analysis. *Computational Statistics & Data Analysis*, 55(1):199–212.

Meinshausen, N., Bühlmann, P., et al. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

Moura, G. V., Santos, A. A., and Ruiz, E. (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking & Finance*.

Nascimento, V. H. and Zakharov, Y. V. (2016). RLS adaptive filter with inequality constraints. *IEEE Signal Processing Letters*, 23:752–756.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

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

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics*, 211:388–413.

Petersen, K. B. and Pedersen, M. S. (2012). The matrix cookbook. Available at http://www.math.uwaterloo.ca/ hwolkowi/matrixcookbook.pdf.

Raftery, A. E., Kárný, M., and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52:52–66.

Reh, L., Krüger, F., and Liesenfeld, R. (2021). Predicting the global minimum variance portfolio. *Available at SSRN 3471216*.

Rue, H. (2001). Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338.

Schön, T., Gustafsson, F., and Nordlund, P.-J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on signal processing*, 53(7):2279–2289.

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

Shin, H.-S. and Lee, H.-I. (2020). A new exponential forgetting algorithm for recursive least-squares parameter estimation. Working paper, available at https://arxiv.org/abs/2004.03910v1.

Sims, C. A. (2001). Evolving post-world war II US inflation dynamics: Comment. *NBER macroeconomics annual*, 16:373–379.

Stock, J. H. (2001). Evolving post-world war II US inflation dynamics: Comment. *NBER macroeconomics annual*, 16:379–387.

Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79. 1.

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

Uhlig, H. (1994). On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, pages 395–405.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, 65:59–74.

Voev, V. (2009). On the economic evaluation of volatility forecasts. Working Paper, CRE-ATES (Research Paper 2009-56).

Wang, L., Chen, Z., Wang, C. D., and Li, R. (2020). Ultrahigh dimensional precision matrix estimation via refitted cross validation. *Journal of Econometrics*, 215(1):118–130.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21:1455–1508.

Windle, J. and Carvalho, C. M. (2014). A tractable state-space model for symmetric positive-definite matrices. *Bayesian Analysis*, 9(4):759–792.

Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534.

Yen, Y.-M. (2016). Sparse weighted-norm minimum variance portfolios. *Review of Finance*, 20(3):1259–1287.

Young, P. C. (2011). *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Springer.

Zhao, Z., Ledoit, O., and Jiang, H. (2021). Risk reduction and efficiency increase in large portfolios: Gross-exposure constraints and shrinkage of the covariance matrix. *Journal of Financial Econometrics*.