CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

No. 582

Nikolaus Hautsch and Stefan Voigt

Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

The CFS Working Paper Series

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty*

Nikolaus Hautsch

University of Vienna, CFS,
Vienna Graduate School of Finance

Stefan Voigt

Vienna Graduate School of Finance

Abstract

We theoretically and empirically study large-scale portfolio allocation problems when transaction costs are taken into account in the optimization problem. We show that transaction costs act on the one hand as a turnover penalization and on the other hand as a regularization, which shrinks the covariance matrix. As an empirical framework, we propose a flexible econometric setting for portfolio optimization under transaction costs, which incorporates parameter uncertainty and combines predictive distributions of individual models using optimal prediction pooling. We consider predictive distributions resulting from high-frequency based covariance matrix estimates, daily stochastic volatility factor models and regularized rolling window covariance estimates, among others. Using data capturing several hundred Nasdaq stocks over more than 10 years, we illustrate that transaction cost regularization (even to small extent) is crucial in order to produce allocations with positive Sharpe ratios. We moreover show that performance differences between individual models decline when transaction costs are considered. Nevertheless, it turns out that adaptive mixtures based on high-frequency and low-frequency information yield the highest performance. Portfolio bootstrap reveals that naive $1/N$-allocations and global minimum variance allocations (with and without short sales constraints) are significantly outperformed in terms of Sharpe ratios and utility gains.

*Nikolaus Hautsch (nikolaus.hautsch@univie.ac.at), Department of Statistics and Operations Research, Faculty of Business, Economics and Statistics, University of Vienna, as well as Vienna Graduate School of Finance (VGSF) and Center for Financial Studies (CFS), Frankfurt. Stefan Voigt (stefan.voigt@vgsf.ac.at), Vienna Graduate School of Finance (VGSF). We thank Gregor Kastner, Allan Timmermann and participants of the Vienna-Copenhagen Conference on Financial Econometrics, 2017, the 3rd Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance, the Conference on Big Data in Predictive Dynamic Econometric Modeling, Pennsylvania, the Conference on Stochastic Dynamical Models in Mathematical Finance, Econometrics, and Actuarial Sciences, Lausanne, 2017, the 10th annual SoFiE conference, New York, the FMA European Conference, Lisbon, the 70th European Meeting of the Econometric Society, Lisbon, the 4th annual conference of the International Association for Applied Econometrics, Sapporo, and the Annual Conference 2017 of the German Economic Association for valuable feedback. Moreover, we greatly acknowledge the use of computing resources by the Vienna Scientific Cluster.

JEL classification: C58, C52, C11, G11 **Keywords:** Portfolio choice, transaction costs, model uncertainty, regularization, high frequency data

1 Introduction

Optimizing large-scale portfolio allocations remains a challenge for econometricians and practitioners due to (i) the noisiness of parameter estimates in large dimensions, (ii) model uncertainty and time variations in individual models' forecasting performance, and (iii) the presence of transaction costs, making otherwise optimal turnover strategies costly and thus sub-optimal.

Though there is a huge literature on the statistics of portfolio allocation, the literature is widely fragmented and typically only focuses on partial aspects. For instance, a substantial part of the literature concentrates on the problem of estimating vast-dimensional covariance matrices by means of regularization techniques, see, e.g., Ledoit and Wolf (2003), Ledoit and Wolf (2004), Fan et al. (2008) and Ledoit and Wolf (2012), among others. This literature has been boosted by the availability of high-frequency (HF) data which opens an additional channel to increase the precision of covariance estimates and forecasts.[1] Another segment of the literature studies the effects of ignoring parameter uncertainty and model uncertainty arising from changing market regimes and structural breaks.[2]

Further literature is devoted to the role of transaction costs in portfolio allocation strategies. In fact, in the presence of transaction costs, the benefits of re-allocating wealth may be smaller than the costs associated with turnover.[3] While this aspect is studied theoretically in mathematical finance, see, e.g., Davis and Norman (1990), there are only very few empirical ap-

[1]The benefits of high-frequency data to estimate covariances have been documented by a wide range of studies, e.g., Andersen and Bollerslev (1998), Andersen et al. (2001), Barndorff-Nielsen (2002), Barndorff-Nielsen and Shephard (2004).

[2]The effect of ignoring estimation uncertainty is considered, among others, by Jobson et al. (1979), Jorion (1986), Chopra and Ziemba (1993), Uppal and Wang (2003) and DeMiguel et al. (2009). Model uncertainty is investigated, for instance, by Wang (2005), Garlappi et al. (2007) and Pflug et al. (2012).

[3]See Brandt et al. (2009) for an excellent review on common pitfalls in portfolio optimization.

proaches, which explicitly account for transaction costs, see, for instance, Acharya and Pedersen (2005) and Gârleanu (2009).

In fact, in most empirical studies, transaction costs are incorporated *ex post* by analyzing to which extent a certain portfolio strategy would have survived in the presence of transaction costs of given size, see, e.g., Hautsch et al. (2015) or Bollerslev et al. (2016). In financial practice, however, the costs of portfolio re-balancing are taken into account *ex ante* and thus are part of the optimization problem.

The objective of this paper is to theoretically and empirically analyze large-dimensional portfolio allocation problems under transaction costs and model uncertainty. Our contribution is twofold: On the one hand, we show that the explicit inclusion of transaction costs in the optimization problem implies a regularization of the estimation problem. In particular, *quadratic* transaction costs can be interpreted as shrinkage towards a diagonal matrix, implying a trade-off between transaction costs and potential diversification benefits. Transaction costs *proportional* to the amount of re-balancing imply a regularization of the variance-covariance, which acts similarly as the least absolute shrinkage and selection operator (Lasso) by Tibshirani (1996) in a regression problem. It thus implies to put more weight on a buy-and-hold strategy. The regulatory effect of transaction costs results in better conditioned covariance estimates and moreover implies a turnover penalization, which significantly reduces the amount (and frequency) of re-balancing. These mechanisms imply strong improvements of portfolio allocations in terms of Sharpe ratios or utility-based measures compared to the case where transaction costs are neglected.

On the other hand, we perform a reality check by empirically analyzing the role of transaction costs in a high-dimensional and preferably realistic setting. We take the perspective of an investor who is monitoring the portfolio allocation on a daily basis while accounting for the (expected) costs of re-balancing. The underlying portfolio optimization setting accounts for parameter uncertainty and model uncertainty, while utilizing not only predictions of the covariance structure but also of higher-order moments of the asset return distribution. Model uncertainty is taken into account by considering time-varying combinations of predictive distributions resulting from competing models using optimal prediction pooling according to Geweke et al. (2011). In this way, we analyze to which extent the predictive ability of individual models changes over time and how suitable forecast combinations may result into better portfolio allocations.

We are particularly interested in the question whether the power of HF data for global minimum variance (GMV) allocations, as recently documented by Liu (2009), Hautsch et al. (2015) and Lunde et al. (2016), among others, still pays out in such a framework. The high responsiveness of HF-based predictions is particularly beneficial in high-volatility market regimes, but may create (too) high turnover. Our framework allows us to analyze to which extent HF data are still useful when transaction costs are explicitly taken into account.

The downside of such generality is that the underlying optimization problem cannot be solved in closed form and requires (high-dimensional) numerical integration. We therefore pose the econometric model in a Bayesian framework, which allows us to integrate out parameter uncertainty and to construct posterior predictive asset return distributions based on time-varying mixtures. Optimality of the portfolio weights is ensured with respect to the predicted out-of-sample utility net of transaction costs. The entire setup is complemented by a portfolio bootstrap by performing the analysis based on randomized sub-samples out of the underlying asset universe. In this way, we are able to gain insights into the statistical significance of various portfolio performance measures.

We analyze a large-scale setting based on all constituents of the S&P500 index, which are continuously traded on Nasdaq between 2007 and 2017, corresponding to 308 stocks. Forecasts of the daily asset return distribution are produced based on three major model classes. On the one hand, utilizing HF message data, we compute estimates of daily asset return covariance matrices using blocked realized kernels according to Hautsch et al. (2012). The kernel estimates are equipped with a Gaussian-Wishart mixture in the spirit of Jin and Maheu (2013) to capture the entire return distribution. Moreover, we compute predictive distributions resulting from a daily multivariate stochastic volatility factor model in the spirit of Chib et al. (2006). Due to recent advances in the development of numerically efficient simulation techniques by Kastner et al. (2017), it is possible to estimate such a model based on high-dimensional data and makes it one of the very few sufficiently flexible parametric models for return distributions covering several hundreds of assets.[4] As a third model class, representing traditional estimators based on rolling windows, we utilize the sample covariance and the (linear) shrinkage estimator proposed by Ledoit and Wolf (2004).

To our best knowledge, this paper provides the first study evaluating the predictive power of state-of-the-art high-frequency and "low-frequency" models in a large-scale portfolio framework under such generality, while utilizing data of 2409 trading days and more than 73 Billion high-frequency observations. Our approach brings together concepts from (i) Bayesian estimation for portfolio optimization, ii) regularization and turnover penalization, (iii) predictive model combinations in high dimensions and (iv) HF-based covariance modeling and prediction.[5]

We can summarize the following findings: First, none of the underlying (state-of-the-art) predictive models is able to produce positive Sharpe ratios when transaction costs are *not* taken into account ex ante. This is mainly due to high turnover implied by (too) frequent re-balancing. Second, when incorporating transaction costs into the optimization problem, performance differences between competing predictive models for the return distribution become smaller than in the case without an explicit inclusion of transaction costs. It is shown that none of the under-

[4]So far, stochastic volatility models have been shown to be beneficial in portfolio allocation by Aguilar and West (2000) and Han (2006) for just up to 20 assets.

[5]Bayesian estimation for portfolio optimization has been applied within a wide range of applications, starting with Brown (1976) and Jorion (1986). Imposing turnover penalties is related to the ideas of Brodie et al. (2009) and Gârleanu and Pedersen (2013). Tu and Zhou (2010), Tu and Zhou (2011) and Anderson and Cheng (2016) emphasis the benefits of model combination in portfolio decision theory. Sequential learning in a two-dimensional asset horizon is performed by Johannes et al. (2014). However, none of these approaches is focusing on mixtures of HF and lower frequencies approaches and aims at large dimensional allocation problems.

lying approaches does produce significant utility gains on top of each other. We thus conclude that the respective pros and cons of the individual models in terms of efficiency, predictive accuracy and stability of covariance estimates are leveled out under turnover regularization. We moreover conclude that the benefits of HF-based covariance predictions are smaller than in case of daily GMV allocations.

Third, despite of a similar performance of individual predictive models, mixing high-frequency and low-frequency information is beneficial and yields significantly higher Sharpe ratios. This is due to time variations in the individual model's predictive ability. Taking this into account when constructing (time-varying) combination weights, the relative contribution of HF-based return predictions is on average approximately 40%. The remaining 60% are provided by both the multivariate stochastic volatility model (approximately 30%) and predictions based on the shrunken sample covariances. HF-based predictions are particularly superior in volatile market periods, but are dominated by SV-based predictions in more calm periods. Fourth, naive strategies or GMV strategies are statistically and economically significantly outperformed.

The structure of this paper is as follows: Section 2 theoretically studies the effect of transaction costs on the optimal portfolio structure. Section 3 gives the econometric setup accounting for parameter and model uncertainty. Section 4 presents the underlying predictive models. In Section 5, we describe the data and present the empirical results. Finally, Section 6 concludes.

2 The Role of Transaction Costs

2.1 Decision Framework

We consider an investor equipped with power utility function $U_\gamma(r) := \frac{r^{1-\gamma}}{1-\gamma}$ depending on return r and risk aversion parameter $\gamma > 1$. At every period t, the investor allocates her wealth among N distinct risky assets with the aim to maximize expected utility at $t+1$ by choosing the allocation vector $\omega_{t+1} \in \mathbb{R}^N$. We impose the constraint $\sum_{i=1}^{N} \omega_{t+1,i} = 1$.

The choice of ω_{t+1} is based on drawing inference from observed data. The information set at time t consists of the time series of past returns $R_t = (r_1', \ldots, r_t')' \in \mathbb{R}^{t \times N}$, where $r_t := \frac{p_t - p_{t-1}}{p_{t-1}} \in \mathbb{R}^N$ are the gross returns computed using end-of-day asset prices $p_t := (p_{t,1}, \ldots, p_{t,N}) \in \mathbb{R}^N$. The set of information may contain additional variables $\mathcal{H}_t$, as, e.g., intra-day data.

We define an optimal portfolio as the allocation, which maximizes expected utility of the investor after subtracting transaction costs arising from re-balancing. We denote transaction costs as $\nu_t(\omega)$, depending on the desired portfolio weight ω and reflecting broker fees and implementation shortfall.

Transaction costs are a function of the distance between the new allocation ω_{t+1} and the allocation right before readjustment, $\omega_{t^+} := \frac{\omega_t \circ r_t}{\iota'(\omega_t \circ r_t)}$, where ι is a vector of ones and the operator $\circ$ denotes element-wise multiplication. The vector ω_{t^+} builds on allocation ω_t, which has been considered as being optimal given expectations in $t-1$, but effectively changed due to returns between $t-1$ and t.

At time t, the investor monitors her portfolio and solves a static maximization problem conditional on the current beliefs on the distribution of returns $p_t\left(r_{t+1}|\mathcal{D}\right) := p(r_{t+1}|\mathcal{D}, R_t, \mathcal{H}_t)$ of the next period and the current portfolio weights ω_{t^+}:

$$\begin{aligned} \omega_{t+1}^* &:= \underset{\omega \in \mathbb{R}^N,\ \iota'\omega = 1}{\arg\max}\ E\left(U_\gamma\left(\omega' r_{t+1} - \nu_t(\omega)\right)|\mathcal{D}, R_t, \mathcal{H}_t\right) \\ &= \underset{\omega \in \mathbb{R}^N,\ \iota'\omega = 1}{\arg\max} \int_{\mathbb{R}^N} U_\gamma\left(\omega' r_{t+1} - \nu_t(\omega)\right) p_t(r_{t+1}|\mathcal{D}) \mathrm{d}r_{t+1}. \end{aligned} \tag{EU}$$

Note that optimization problem (EU) reflects the problem of an investor who constantly monitors her portfolio and exploits all available information, but re-balances only if the costs implied by deviations from the path of optimal allocations exceed the costs of re-balancing. This form of myopic portfolio optimization ensures optimality (after transaction costs) of allocations at each point in time. Accordingly, the optimal wealth allocation ω^*_{t+1} from representation (EU) is governed by i) the structure of turnover penalization $\nu_t(\omega)$, and ii) the return forecasts $p_t(r_{t+1}|\mathcal{D})$.

2.2 Transaction Costs in Case of Gaussian Returns

In general, the solution to the optimization problem (EU) cannot be derived analytically but needs to be approximated using numerical methods. However, assuming $p_t(r_{t+1}|\mathcal{D})$ being a multivariate log-normal density with known parameters Σ and μ, problem (EU) coincides with the initial Markowitz (1952) approach and yields an analytical solution, resulting from the maximization of the certainty equivalent (CE) after transaction costs,

$$\omega^*_{t+1} = \underset{\omega \in \mathbb{R}^N, \iota'\omega=1}{\arg\max}\ \omega'\mu - \nu_t(\omega) - \frac{\gamma}{2}\omega'\Sigma\omega. \tag{1}$$

2.2.1 Quadratic transaction costs

We model the transaction costs $\nu_t(\omega_{t+1}) := \nu(\omega_{t+1}, \omega_{t^+}, \beta)$ for shifting wealth from allocation ω_{t^+} to ω_{t+1} as a quadratic function given by

$$\nu(\omega_{t+1}, \omega_{t^+}, \beta) = \frac{\beta}{2}(\omega_{t+1} - \omega_{t^+})'(\omega_{t+1} - \omega_{t^+}), \tag{2}$$

with cost parameter $\beta > 0$. The allocation ω^*_{t+1} according to (1) can then be restated as

$$\begin{aligned} \omega^*_{t+1} &= \underset{\omega \in \mathbb{R}^N, \iota'\omega=1}{\arg\max}\ \omega'\mu - \frac{\beta}{2}(\omega_{t+1} - \omega_{t^+})'(\omega_{t+1} - \omega_{t^+}) - \frac{\gamma}{2}\omega'\Sigma\omega \\ &= \underset{\omega \in \mathbb{R}^N,\ \iota'\omega=1}{\arg\min}\ \frac{\gamma}{2}\omega'\Sigma^*\omega - \omega'\mu^*, \end{aligned} \tag{3}$$

with

$$\Sigma^* := \frac{\beta}{\gamma} I + \Sigma, \tag{4}$$

$$\mu^* := \mu + \beta \omega_{t^+}, \tag{5}$$

where I denotes the identity matrix. The optimization problem with quadratic transaction costs can be thus interpreted as a classical mean-variance problem *without* transaction costs, where (i) the covariance matrix Σ is regularized towards the identity matrix (with $\frac{\beta}{\gamma}$ serving as shrinkage parameter) and the mean is shifted by $\beta\omega_{t^+}$. Hence, if β increases, Σ^* is shifted towards a diagonal matrix representing the case of uncorrelated assets. Higher transaction costs are therefore equivalent to a setup with less diversification benefits. The shift from μ to $\mu^* = \mu + \beta\omega_{t^+}$ can be alternatively interpreted by exploiting $\omega'\iota = 1$ and reformulating the problem as

$$\omega^*_{t+1} = \underset{\omega \in \mathbb{R}^N,\ \iota'\omega = 1}{\arg\min} \ \frac{\gamma}{2}\omega'\Sigma^*\omega - \omega'\left(\mu + \beta\left(\omega_{t^+} - \frac{1}{N}\iota\right)\right).$$

From this representation it becomes obvious that the shift of the mean vector is proportional to the deviations of the current allocation to the $1/N$-setting. This can be interpreted as putting more weight on assets with (already) high exposure. Proposition 1 shows the effect of rising transaction costs on optimal re-balancing.

Proposition 1.

$$\lim_{\beta \to \infty} \omega^*_{t+1} = \left(I - \frac{1}{N}\iota\iota'\right)\omega_{t^+} + \frac{1}{N}\iota = \omega_{t^+}. \tag{6}$$

Proof. See Appendix A. □

Hence, if the transaction costs are prohibitively large, the investor may not implement the efficient portfolio despite her knowledge of the true return parameters μ and Σ. The effect of transaction costs in the long-run can be analyzed in more depth by considering the well-known representation of the mean-variance efficient portfolio,

$$\omega(\mu, \Sigma) := \frac{1}{\gamma}\left(\underbrace{\Sigma^{-1} - \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota\iota'\Sigma^{-1}}_{:=A(\Sigma)}\right)\mu + \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota. \tag{7}$$

If ω_0 denotes the initial allocation, sequential re-balancing allows us to study the long-run effect, given by

$$\omega_T = \sum_{i=0}^{T-1}\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^i \omega(\mu, \Sigma^*) + \left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^T \omega_0. \tag{8}$$

Hence, ω_T can be interpreted as a weighted average of $\omega(\mu, \Sigma^*)$ and the initial allocation ω_0, where the weights depend on the ratio β/γ. The following proposition shows, however, that a range for β exists (with critical upper threshold), for which the initial allocation can be ignored in the long-run.

Proposition 2. $\exists\beta^* > 0\ \forall\tilde{\beta} \in [0, \beta^*) : \left\|\left(\frac{\tilde{\beta}}{\gamma}A(\Sigma^*)\right)\right\|_F < 1$, *where* $\|\cdot\|_F$ *denotes the Frobenius norm* $\|A\|_F := \sqrt{\sum_{i=1}^{N}\sum_{i=1}^{N} a_{i,j}^2}$.

Proof. See Appendix A. □

Using Proposition 2 for $T \to \infty$ and $\beta < \beta^*$, the series $\sum_{i=0}^{T}\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^i$ converges to $\left(I - \frac{\beta}{\gamma}A(\Sigma^*)\right)^{-1}$ and $\lim_{i\to\infty}\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^i = 0$. In the long-run, we then obtain

$$\omega_\infty = \left(I - \frac{\beta}{\gamma}A(\Sigma^*)\right)^{-1}\omega(\mu, \Sigma^*).$$

Note, that the location of the initial portfolio ω_0 itself does not play a role on the upper threshold β^* which ensures the long-run convergence towards ω_∞. Instead, β^* is affected only by the risk aversion γ and the eigenvalues of Σ.

2.2.2 Proportional (L_1) transaction costs

Although attractive from an analytical perspective, transaction costs of quadratic form may represent an unrealistic proxy of costs associated with trading in real financial markets. Instead, in literature there is widespread use of transaction cost measures proportional to the sum of absolute re-balancing (L_1-norm of re-balancing), which impose stronger penalization on turnover and arguably are more realistic, see, for instance DeMiguel et al. (2009). Transaction costs proportional to the L_1 norm of $\omega_{t+1} - \omega_{t^+}$ yield the form

$$\nu\left(\omega_{t+1}, \omega_{t^+}, \beta\right) = \beta ||\omega_{t+1} - \omega_{t^+}||_1 := \beta \sum_{i=1}^{N} |\omega_{t+1,i} - \omega_{t^+,i}|, \tag{9}$$

with cost parameter $\beta > 0$. Although the effect of L_1 transaction costs on the optimal portfolio cannot be derived in closed-form comparable to the quadratic (L_2) case, the impact of turnover penalization can still be interpreted as a form of regularization. If we assume for simplicity of illustration $\mu = 0$, the optimization problem (1) corresponds to

$$\omega_{t+1}^* := \underset{\omega \in \mathbb{R}^N,\ \iota'\omega = 1}{\arg\min} \frac{\gamma}{2}\omega'\Sigma\omega + \beta||\omega_{t+1} - \omega_{t^+}||_1 \tag{10}$$

$$= \underset{\Delta \in \mathbb{R}^N,\ \iota'\Delta = 0}{\arg\min} \gamma\Delta'\Sigma\omega_{t^+} + \frac{\gamma}{2}\Delta'\Sigma\Delta + \beta||\Delta||_1. \tag{11}$$

The first-order conditions for the constrained optimization are

$$\gamma\Sigma \underbrace{\left(\Delta + \omega_{t^+}\right)}_{\omega_{t+1}^*} + \beta\tilde{g} - \lambda\iota = 0, \tag{12}$$

$$\Delta'\iota = 0, \tag{13}$$

where $\tilde{g}$ is the vector of sub-derivatives of $||\omega_{t+1} - \omega_{t^+}||_1$, i.e., $\tilde{g} := \partial||\omega_{t+1} - \omega_{t^+}||_1/\partial\omega_{t+1}$, consisting of elements which are 1 or -1 in case $\omega_{t+1,i} - \omega_{t^+,i} > 0$ or $\omega_{t+1,i} - \omega_{t^+,i} < 0$, respectively, or $\in [-1, 1]$ in case $\omega_{t+1,i} - \omega_{t^+,i} = 0$. Solving for ω^*_{t+1} yields

$$\omega^*_{t+1} = \left(1 + \frac{\beta}{\gamma}\iota'\Sigma^{-1}\tilde{g}\right)\omega_{\text{mvp}} - \frac{\beta}{\gamma}\Sigma^{-1}\tilde{g}, \tag{14}$$

where $\omega_{\text{mvp}} := \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota$ corresponds to the weights of the GMV portfolio. Proposition 3 shows that this optimization problem can be formulated as a (regularized) minimum variance problem.

Proposition 3. *Portfolio optimization problem (10) is equivalent to the minimum variance problem with*

$$\omega^*_{t+1} = \underset{\omega\in\mathbb{R}^N,\ \iota'\omega=1}{\arg\min}\ \omega'\Sigma_{\frac{\beta}{\gamma}}\omega, \tag{15}$$

where $\Sigma_{\frac{\beta}{\gamma}} = \Sigma + \frac{\beta}{\gamma}(g^*\iota' + \iota g^*)$, *and* g^* *is the subgradient of* $||\omega^*_{t+1} - \omega_{t^+}||_1$.

Proof. See Appendix A. □

The interpretation of this result is straightforward: the effect of imposing transaction costs proportional to the L_1 norm of re-balancing can be interpreted as a standard GMV setup with a regularized version of the variance-covariance matrix. The form of the matrix $\Sigma_{\frac{\beta}{\gamma}}$ implies that for high transaction costs β, more weight is put on those pairs of assets, whose exposure is re-balanced in the same direction. The result is similar to Fan et al. (2012), who show that the risk minimization problem with constrained weights

$$\omega^*_{t+1} = \underset{\omega\in\mathbb{R}^N,\ \iota'\omega=1,\ ||\omega||_1\leq\vartheta}{\arg\min}\ \omega'\Sigma\omega \tag{16}$$

can be interpreted as the minimum variance problem

$$\omega^*_{t+1} = \underset{\omega \in \mathbb{R}^N,\ \iota'\omega=1}{\arg\min}\ \omega'\tilde{\Sigma}\omega, \tag{17}$$

with $\tilde{\Sigma} = \Sigma + \lambda\left(g\iota' + \iota g'\right)$, where λ is a Lagrange multiplier and g is the subgradient vector of the function $||\omega||_1$ evaluated at the the solution of optimization problem (16). Note, however, that the transaction cost parameter β is given to the investor, whereas ϑ is an endogenously imposed restriction with the aim to decrease the impact of estimation error.

Investigating the long-run effect of the initial portfolio ω_0 in the presence of L_1 transaction costs in the spirit of Proposition 1 is complex as analytically tractable representations are not easily available. General insights from the L_2 benchmark case, however, can be transferred to the setup with L_1 transaction costs: First, a high cost parameter β may prevent the investor to implement the efficient portfolio. Second, as the L_2 norm of any vector is bounded from above by its L_1 norm, L_1 penalization is always stronger than in the case of quadratic transaction costs. Therefore, we expect that the convergence of portfolios from the initial allocation ω_0 towards the efficient portfolio is generally slower, but qualitatively similar.

2.2.3 Empirical implications

To empirically illustrate the effects discussed above, we compute the performance of portfolios after transaction costs based on $N = 308$ assets and daily readjustments based on data ranging from June 2007 to March 2017.[6] The unknown variance-covariance matrix Σ_t is estimated in two ways: We compute the sample variance-covariance estimator S_t and the shrinkage estimator $\hat{\Sigma}^{t,\text{Shrink}}$ by Ledoit and Wolf (2004) with a rolling window of length $h = 500$ days. We refrain from estimating the mean and set $\mu_t = 0$. The initial portfolios weights are set to $\frac{1}{N}\iota$, corresponding to the naive portfolio. Then, for a fixed β and daily estimates of Σ_t,

[6] A description of the dataset and the underlying estimators is given in more detail in Section 5.

portfolio weights are re-balanced as solutions of optimization problem (3) using $\gamma = 4$. This yields a time series of optimal portfolios ω_t^β and realized portfolio returns $r_t^\beta := r_t'\omega_t^\beta$. Subtracting transaction costs then yields the realized portfolio returns net of transaction costs $r_t^{\beta,\text{nTC}} := r_t^\beta - \beta\left(\omega_{t+1}^\beta - \omega_{t^+}^\beta\right)'\left(\omega_{t+1}^\beta - \omega_{t^+}^\beta\right)$.

Figure 1 displays annualized Sharpe ratios[7] after subtracting transaction costs for both the sample covariance and the shrinkage estimator for a range of different values of β, measured in basis points. The purple line represents Sharpe ratios implied by the naive allocation.

We observe that the naive portfolio is clearly outperformed. This is remarkable, as it is well-known that parameter uncertainty especially in high dimensions often leads to superior performance of the naive allocation (see, e.g. DeMiguel et al. (2009)). We moreover find that already very small positive values of β have a strong regulatory effect on the covariance matrix. Recall that quadratic transaction costs directly affect the diagonal elements and thus the eigenvalues of Σ_t. Inflating then each of the 308 eigenvalues already by a small magnitude has a substantial effect on the conditioning of the covariance matrix. We observe that transaction costs of just 1 bp significantly increase the conditioning number and strongly stabilize S_t and particularly its inverse, S_t^{-1}. In fact, the optimal portfolio based on the sample variance-covariance matrix which adjusts ex-ante for transaction costs of $\beta = 1bp$ leads to a net-of-transaction-costs Sharpe ratio of 0.37, whereas neglecting transaction costs in the optimization yields a Sharpe ratio of only 0.12. This effect is to a large extent a pure regularization effect. For rising values of β, this effect marginally declines and leads to a declining performance for values of β between 10 and 50bp. We therefore observe a dual role of transaction costs. On the one hand, they improve the conditioning of the covariance matrix by inflating the eigenvalues. On the other hand, they reduce the mean portfolio return. Both effects influence the Sharpe ratio in opposite direction causing the concavity of graphs for values of β up to approximately 50bp.

[7]Computed as the ratio of the annualized sample mean and standard deviation of $r_t^{\beta,\text{nTC}}$.

For higher values of β we observe a similar pattern: Here, a decline in re-balancing costs due to the implied turnover penalization kicks in and implies an increase of the Sharpe ratio. If the cost parameter β, however, gets too high, the adverse effect of transaction costs on the portfolio's expected returns dominate. Hence, as predicted by Proposition 1, the allocation is ultimately pushed back to the initial portfolio (the $1/N$ portfolio in this case). This is reflected by the lower panel of Figure 1, showing the distance to the initial allocation in terms of the average L_1 distance.

Moreover, the described effects are much more pronounced for the sample covariance than for its shrunken counterpart. As the latter is already regularized, additional regulation implied by turnover penalization obviously has a lower impact. Nevertheless, turnover regularization implies that forecasts even based on the sample covariance generate reasonable Sharpe ratios, which tend to perform equally well than those implied by a (linear) shrinkage estimator. Hence, we observe that differences in the out-of-sample performance between the two approaches decline if the turnover regularization becomes stronger.

Figure 2 visualizes the effect of L_1 transaction costs. Note that L_1 transaction costs imply a regulation which acts similarly to a Lasso penalization. Such a penalization is less smooth than that implied by quadratic transaction costs and implies a stronger dependence of the portfolio weights on the previous day's allocation. This strong path-dependence implies that the plots in Figure 2 are less smooth than in the case of quadratic transaction costs. This affects our evaluation, as the paths of the portfolio weights may differ substantially over time if the cost parameter β is slightly changed.

However, we find similar effects as for quadratic transaction costs. For low values of β, the Sharpe ratio increases in β. Here, the effects of covariance regularization and reduction of turnover overcompensate the effect of declining portfolio returns (after transaction costs). For larger values of β, we observe the expected convergence towards the performance of the naive portfolio. For this range of β, the two underlying approaches perform rather en par. Hence, additional regularization as implied by the shrinkage estimator is not (very) effective and does not generate excess returns on top of the sample variance-covariance matrix.

3 Basic Econometric Setup

3.1 Parameter Uncertainty and Model Combination

The optimization problem (EU) posses the challenge of providing a sensible density $p_t(r_{t+1}|\mathcal{D})$ of future returns. The predictive density should reflect dynamics of the return distribution in a suitable way, which opens many different dimensions on how to choose a model $\mathcal{M}_k$. The model $\mathcal{M}_k$ reflects assumptions regarding the return generating process in form of a likelihood function $\mathcal{L}\left(r_t|\Theta, \mathcal{H}_t, \mathcal{M}_k\right)$ depending on unknown parameters Θ. Assuming that future returns are distributed as $\mathcal{L}\left(r_t|\hat{\Theta}, \mathcal{H}_t, \mathcal{M}_k\right)$, where $\hat{\Theta}$ is a point estimate of the parameters Θ, however, would imply that the uncertainty perceived by the investor ignores estimation error.[8] Consequently, the resulting portfolio weights would be sub-optimal and the exposure of the investor to risky assets would be unreasonably high.

To accommodate such parameter uncertainty and to pose a setting, where the optimization problem (EU) can be naturally addressed by numerical integration techniques, we employ a Bayesian approach. Hence, by defining a model $\mathcal{M}_k$ implying the likelihood $\mathcal{L}\left(r_t|\Theta, \mathcal{H}_t, \mathcal{M}_k\right)$ and choosing a prior distribution $\pi(\Theta)$, the posterior distribution

$$\pi(\Theta|R_t, \mathcal{H}_t) \propto \mathcal{L}(R_t|\Theta, \mathcal{H}_t, \mathcal{M}_k)\pi(\Theta) \tag{18}$$

reflects beliefs about the distribution of the parameters after observing the set of available

[8] See e.g. Brown (1976), Kan and Zhou (2007) or Avramov and Zhou (2010) for detailed treatments of the impact of parameter uncertainty on optimal portfolios.

information, $(R_t, \mathcal{H}_t)$. The (posterior) predictive distribution of the returns is then given by

$$r_{\mathcal{M}_k,t+1} \sim p(r_{t+1}|R_t, \mathcal{H}_t, \mathcal{M}_k) := \int \mathcal{L}(r_{t+1}|\Theta, \mathcal{H}_t, \mathcal{M}_k)\pi(\Theta|R_t, \mathcal{H}_t)d\Theta. \tag{19}$$

If the parameters cannot be estimated with high precision, the posterior distribution $\pi(\Theta|R_t, \mathcal{H}_t)$ yields a predictive return distribution with more mass in the tails than focusing only on $\mathcal{L}(r_{t+1}|\hat{\Theta}, \mathcal{H}_t, \mathcal{M}_k)$. Therefore, parameter uncertainty automatically leads to a higher predicted probability of tail events, implying the fat-tailedness of the posterior predictive asset return distribution.

Moreover, potential structural changes in the return distribution and time-varying parameters make it hard to identify a single predictive model which consistently outperforms all other models. Therefore, an investor may instead combine predictions of K distinct predictive models $\mathcal{D} := \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$, reflecting either personal preferences, data availability or theoretical considerations.[9] Stacking the predictive distributions yields

$$r^{\text{vec}}_{\mathcal{D},t+1} := \text{vec}\left(\{r_{\mathcal{M}_1,t+1}, \ldots, r_{\mathcal{M}_K,t+1}\}\right) \in \mathbb{R}^{NK\times 1}. \tag{20}$$

The joint predictive distribution $p_t(r_{t+1}|\mathcal{D})$ is computed conditional on combination weights $c_t \in \mathbb{R}^K$, which can be interpreted as discrete probabilities over the set of models $\mathcal{D}$. The probabilistic interpretation of the combination scheme is justified by enforcing that all weights take positive values and add up to one:

$$c_t \in \Delta_{[0,1]^K} := \left\{c \in \mathbb{R}^K : c_i \geq 0 \ \forall i = 1, \ldots, K \text{ and } \sum_{i=1}^{K} c_i = 1\right\}. \tag{21}$$

[9] Model combination in the context of return predictions has a long tradition in econometrics, starting from Bates and Granger (1969). In finance, Avramov (2003), Uppal and Wang (2003), Garlappi et al. (2007), Johannes et al. (2014) and Anderson and Cheng (2016), among others, apply model combinations and investigate the effect of model uncertainty on financial decisions.

This yields the joint predictive distribution

$$p_t(r_{t+1}|\mathcal{D}) := p(r_{t+1}|R_t, \mathcal{H}_t, \mathcal{D}) = \int (c_t \otimes I)' \, p\left(r^{\text{vec}}_{\mathcal{D},t+1}|R_t, \mathcal{H}_t, \mathcal{D}\right) dr^{\text{vec}}_{\mathcal{D},t+1}, \tag{22}$$

corresponding to a mixture distribution with time-varying weights.

Depending on the choice of the combination weights c_t, the scheme balances how much the investment decision is driven by each of the individual models. Well-known approaches to combine different models are, among many others, Bayesian model averaging (Hoeting et al., 1999), optimal prediction pooling (Geweke et al., 2011) and decision-based model combinations (Billio et al., 2013).

We choose c_t conditional on the past data $(R_t, \mathcal{H}_t)$ as a solution to the maximization problem

$$c_t = \underset{c \in \Delta_{[0,1]^K}}{\arg\max} \, f\left(c|R_t, \mathcal{H}_t\right). \tag{23}$$

In line with Geweke et al. (2011) and Durham and Geweke (2014) we focus on evaluating the goodness-of-fit of the predictive distributions as a measure of predictive accuracy based on rolling-window maximization of the predictive log score

$$c_t^* = \underset{c \in \Delta_{[0,1]^K}}{\arg\max} \sum_{i=t-h}^{t} \log \left[\sum_{k=1}^{K} c_k p\left(r_i|R_{i-1}, \mathcal{H}_{i-1}, \mathcal{M}_k\right) \right], \tag{24}$$

where h is the window size and $p\left(r_i|R_{i-1}, \mathcal{H}_{i-1}, \mathcal{M}_k\right)$ is defined as in (19).[10] If the predictive density concentrates around the observed return values, the predictive likelihood is higher. Alternatively, we implemented utility-based model combination in the spirit of Billio et al. (2013) and Pettenuzzo and Ravazzolo (2016) by choosing c_t as a function of past portfolio-performances

[10]In our empirical application we set $h = 250$ days.

net of transaction costs.[11] Though it is appealing to include combination procedures targeted to the objective function of the investor, the combination scheme results in very instable combination weights, putting unit mass in the most successful strategies of the recent past. This results in low portfolio performances due to high turnover and the neglectance of a majority of the information. We therefore refrain from reporting results based on this procedure.

3.2 MCMC-based inference

In general, the objective function of the portfolio optimization problem (EU) is not available in closed form. Furthermore, the posterior predictive distribution may not arise from a well-known class of probability distributions. Therefore, the computation of portfolio weights depends on (Bayesian) computational methods. First, we compute L sample draws $\theta_k^{(1)}, \ldots, \theta_k^{(L)}$ from the posterior distribution $\pi(\Theta_k | R_t, \mathcal{H}_t, \mathcal{M}_k)$ for every model $\mathcal{M}_k \in \mathcal{D}$ using Markov Chain Monte Carlo algorithms.[12] Then, L draws from the posterior predictive distribution are generated by sampling $r_{\mathcal{M}_k,t+1}^{(l)}$ from $\mathcal{L}(r_{t+1} | \theta_k^{(l)}, \mathcal{H}_t, \mathcal{M}_k)$, $\forall l = 1, \ldots, L$. Obtaining the combination weights based on the past predictive performance requires the computation of the predictive scores of every model $\mathcal{M}_k$ as input to the optimization problem (24). This value can be approximated by computing

$$p(r_t^O | R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k) \approx \frac{1}{L} \sum_{l=1}^{L} \mathcal{L}\left(r_t^O | \theta_k^{(l)}, R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k \right), \tag{25}$$

where r_t^O are the observed returns at time t. After updating the combination weights c_t, samples from the joint predictive distribution $p(r_{\mathcal{D},t+1} | R_t, \mathcal{H}_t, \mathcal{D})$ are generated by using the $K \times L$ draws

[11] The combination weights for the utility-based approach can be computed by choosing

$$c_t^U = \underset{c \in \Delta_{[0,1]^K}}{\arg\max} \sum_{i=t-h}^{t} \sum_{k=1}^{K} \left[c_k (\omega_i^{k\prime} r_i - \nu(\omega_i^k, \omega_{i-1^+}^k))^{1-\gamma} \right],$$

where ω_i^k are the optimal portfolio weights at day i of strategy k.

[12] See Hastings (1970) and Chib and Greenberg (1995).

from the posterior models $r^{(l)}_{\mathcal{M}_k,t+1}$. Samples from $r^{(l)}_{\mathcal{D},t+1}$ are obtained by computing

$$r^{(l)}_{\mathcal{D},t+1} = (c_t \otimes I)' \operatorname{vec}\left(\left\{r^{(l)}_{\mathcal{M}_1,t+1}, \ldots, r^{(l)}_{\mathcal{M}_K,t+1}\right\}\right). \tag{26}$$

The integral in optimization problem (EU) can be approximated for any given weight $\omega \in \mathbb{R}^N$ by Monte-Carlo techniques using the draws $r^{(l)}_{\mathcal{D},t+1}$ by computing

$$E\left(\left(\omega' r_{\mathcal{D},t+1} - \nu_t(\omega)\right)^{1-\gamma} | R_t, \mathcal{H}_t, \mathcal{D}\right) \approx \frac{1}{L}\sum_{l=1}^{L}\left(\omega' r^{(l)}_{\mathcal{D},t+1} - \nu_t(\omega)\right)^{1-\gamma}. \tag{27}$$

The vector $\left\{\omega' r^{(l)}_{\mathcal{D},t+1} - \nu_t(\omega)\right\}_{l=1,\ldots,L}$ represents draws from the posterior predictive portfolio return distribution (after subtracting transaction costs) for allocation weight vector ω. The numerical solution $\hat{\omega}_{t+1}$ of the allocation problem (EU) is then obtained by choosing ω such that the sum in (27) is maximized, i.e.,

$$\hat{\omega}_{t+1} := \underset{\omega \in \mathbb{R}^N, \iota'\omega = 1}{\arg\max} \frac{1}{L}\sum_{l=1}^{L}\left(\omega' r^{(l)}_{\mathcal{D},t+1} - \nu_t(\omega)\right)^{1-\gamma}. \tag{28}$$

4 Predictive Models

As predictive models we choose representatives of three major model classes. First, we include covariance forecasts based on *high-frequency* data utilizing blocked realized kernels as proposed by Hautsch et al. (2012). Second, we employ predictions based on parametric models for Σ_t using *daily* data. An approach which is sufficiently flexible, while guaranteeing well-conditioned covariance forecasts, is a stochastic volatility factor model according to Chib et al. (2006). Thanks to the development of numerically efficient simulation techniques by Kastner et al. (2017), (MCMC-based) estimation is tractable even in high dimensions. This makes the model becoming one of the very few parametric models (with sufficient flexibility) which are feasible for data of these dimensions. Third, as a candidate representing the class of shrinkage estimators, we employ the approach by Ledoit and Wolf (2004).

The choice of models is moreover driven by computational tractability in a large-dimensional setting requiring numerical integration through MCMC techniques and in addition a portfolio bootstrap procedure as illustrated in Section 5. We nevertheless believe that these models yield major empirical insights, which can be easily transfered to modified or extended approaches.

4.1 A Wishart Model for Blocked Realized Kernels

Realized measures of volatility based on HF data have been shown to provide accurate estimates of daily covariances.[13] To produce forecasts of covariances based on HF data, we employ blocked realized kernel (BRK) estimates as proposed by Hautsch et al. (2012) to estimate the quadratic variation of the price process based on irregularly spaced and noisy price observations.

The major idea is to estimate the covariance matrix block-wise. Accordingly, stocks are ordered according to their average number of daily mid-quote observations. By separating the stocks into 4 equal-sized groups, the resulting covariance matrix is then decomposed into $b = 10$ blocks representing pair-wise correlations within each group and across groups.[14] We denote the set of indexes of the assets associated with block $b \in 1, \ldots, 10$ by $\mathcal{I}_b$. For each asset i, $\tau_{t,l}^{(i)}$ denotes the time stamp of mid-quote l on day t. The so-called refresh time is the time it takes for all assets in one block to observe at least one mid-quote update and is formally defined as

$$r\tau_{t,1}^{b} := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,1}^{(i)} \right\}, \qquad r\tau_{t,l+1}^{b} := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,N^{(i)}(r\tau_{t,l}^{b})}^{(i)} \right\}, \tag{29}$$

where $N^{(i)}(\tau)$ denotes the number of midquote observations of asset i before time τ. Hence, refresh time sampling synchronizes the data in time with $r\tau_{t,l}^{b}$ denoting the time, where all of

[13] See, e.g., Andersen and Bollerslev (1998), Andersen et al. (2003), Barndorff-Nielsen et al. (2009), Liu (2009) and Hautsch et al. (2015), among others.

[14] Hautsch et al. (2015) find that 4 liquidity groups constitutes a reasonable (data-driven) choice for a similar data set. We implemented the setting for up to 10 groups and find similar results in the given framework.

the assets belonging to group b have been traded at least once since the last refresh time $r\tau^b_{t-1,l}$. Synchronized returns are then obtained as $r^{(i)}_{t,l} := p^{(i)}_{r\tau^b_{t,l}} - p^{(i)}_{r\tau^b_{t,l-1}}$, with $p^{(i)}_{r\tau^b_{t,l}}$ denoting the log mid-quote of asset i at time $r\tau^b_{t,l}$.

Refresh-time-synchronized returns build the basis for the multivariate realized kernel estimator by Barndorff-Nielsen et al. (2011), which allows (under a set of assumptions) to consistently estimate the quadratic covariation of an underlying multivariate Brownian semi-martingale price process which is observed under noise. Applying the multivariate realized kernel on each block of the covariance matrix, we obtain

$$K^b_t := \sum_{h=-L^b_t}^{L^b_t} k\left(\frac{h}{L^b_t + 1}\right) \Gamma^{h,b}_t, \tag{30}$$

where $k(\cdot)$ is the Parzen kernel, $\Gamma^{h,b}_t$ is the h-lag auto-covariance matrix of the assets belonging to block $\mathcal{I}_b$, and L^b_t is a bandwidth parameter, which is optimally chosen according to Barndorff-Nielsen et al. (2011).

The estimates of the correlations between assets in block b take the form

$$\hat{H}^{\mathrm{b}}_t = \left(V^{\mathrm{b}}_t\right)^{-1} K^b_t \left(V^{\mathrm{b}}_t\right)^{-1}, \qquad V^{\mathrm{b}}_t = \mathrm{diag}\left[K^b_t\right]^{1/2}. \tag{31}$$

The blocks $\hat{H}^{\mathrm{b}}_t$ are then stacked as described in Hautsch et al. (2012) to obtain the correlation matrix $\hat{H}_t$.

The diagonal elements of the covariance matrix, $\hat{\sigma}^2_{t,i}$, $i = 1, \ldots, N$, are estimated based on univariate realized kernels according to Barndorff-Nielsen et al. (2008). The resulting variance-covariance matrix is then given by

$$\hat{\Sigma}^{\mathrm{BRK}}_t = \mathrm{diag}\left(\hat{\sigma}^2_{t,1}, \ldots, \hat{\sigma}^2_{t,N}\right)^{1/2} \hat{H}_t \mathrm{diag}\left(\hat{\sigma}^2_{t,1}, \ldots, \hat{\sigma}^2_{t,N}\right)^{1/2}. \tag{32}$$

We stabilize the covariance estimates by smoothing over time and computing simple averages of the last 5 days, i.e., $\hat{\Sigma}_{S,t}^{BRK} := (1/5)\sum_{s=1}^{5}\hat{\Sigma}_{t-s+1}^{BRK}$. A (smoothed) correlation matrix is then obtained as

$$\hat{H}_{S,t} := (V_{S,t})^{-1}\,\hat{\Sigma}_{S,t}^{BRK}\,(V_{S,t})^{-1}\,, \qquad V_{S,t} = \text{diag}\ [K_{S,t}]^{1/2}\,, \tag{33}$$

with $K_{S,t} := 1/S\sum_{s=1}^{S} K_{t-s+1}$. In Section 5, we illustrate that such smoothing of estimates improves the predictive performance of the model as the impact of extreme intra-daily activities is reduced.

To produce not only forecasts of the asset return covariance, but of the entire density, we parametrize a suitable return distribution, which is driven by the dynamics of $\hat{\Sigma}_t^{BRK}$ in the spirit of Jin and Maheu (2013). The dynamics of the predicted return process conditional on the latent covariance Σ_t are modeled as multivariate Gaussian. To capture parameter uncertainty, integrated volatility is modeled as a Wishart distribution.[15] Thus, the model is defined by:

$$\mathcal{L}(r_{t+1}|\Sigma_{t+1}) \sim N(0, \Sigma_{t+1}), \tag{34}$$

$$\Sigma_{t+1}|\kappa, B_t \sim W_N(\kappa, B_t), \tag{35}$$

$$\kappa B_t = \hat{\Sigma}_{S,t}^{\text{BRK}}, \tag{36}$$

$$\kappa \sim \exp{(100)}\, I_{\kappa > N-1}. \tag{37}$$

Although we impose a Gaussian likelihood conditional on Σ_{t+1}, the posterior predictive distribution of the returns exhibit fat tails after marginalizing out Σ_{t+1} due to the choice of the prior.

[15] This represents a multivariate extension of the normal-inverse-gamma approach, applied, for instance, by Barndorff-Nielsen (1997), Andersson (2001), Jensen and Lunde (2001) and Forsberg and Bollerslev (2002).

4.2 Stochastic Volatility Factor Models

Parametric models for return distributions in very high dimensions accommodating time variations in the covariance structure are typically either highly restrictive or computationally (or numerically) not feasible. Even dynamic conditional correlation (DCC) models as proposed by Engle (2002) are not feasible for processes including several hundreds of assets.[16] Likewise, stochastic volatility models allow for flexible (factor) structures but have been computationally not feasible for high-dimensional processes either. Recent advances in MCMC sampling techniques, however, make it possible to estimate stochastic volatility factor models even in very high dimensions while keeping the numerical burden moderate. Employing interweaving schemes to overcome well-known issues of slow convergence and high autocorrelations of MCMC samplers for SV models, Kastner et al. (2017) provide means to considerably boost up computational speed.

We therefore assume a stochastic volatility factor model in the spirit of Shephard (1996), Jacquier et al. (2002) and Chib et al. (2006) as given by

$$r_t = \Lambda V(\xi_t)^{1/2}\zeta_t + Q(\xi_t)^{1/2}\varepsilon_t, \tag{38}$$

where Λ is a $N \times j$ matrix of unknown factor loadings, $Q(\xi_t) = \text{diag}\left(\exp(\xi_{1,t}), \ldots, \exp(\xi_{N,t})\right)$ is a $N \times N$ diagonal matrix of N latent factors capturing idiosyncratic effects, and $V(\xi_t) = \text{diag}\left(\exp(\xi_{N+1,t}), \ldots, \exp(\xi_{N+j,t})\right)$ is a $j \times j$ diagonal matrix of common latent factors. The innovations $\varepsilon_t \in \mathbb{R}^N$ and $\zeta_t \in \mathbb{R}^j$ are assumed to follow independent standard normal distributions. The model thus implies that the covariance matrix of r_t is driven by a factor structure

$$\text{cov}\left(r_t|\xi_t\right) = \Sigma_t(\xi_t) = \Lambda V_t(\xi_t)\Lambda' + Q_t(\xi_t), \tag{39}$$

[16] One notable exception is a shrinkage version of the DCC model as recently proposed by Engle et al. (2017).

with $V_t(\xi_t)$ capturing common factors and $Q_t(\xi_t)$ capturing idiosyncratic components. The covariance elements are thus parametrized in terms of the $N \times j$ unknown parameters, whose dynamics are triggered by j common factors. All $N + j$ latent factors are assumed to follow AR(1) processes,

$$\xi_{it} = \mu_i + \phi_i(\xi_{t-1,i} - \mu_i) + \sigma_i \eta_{t,i} \; i = 1, \ldots, N + j, \tag{40}$$

where the innovations η_t follow independent standard normal distributions and ξ_{i0} is an unknown initial state. The AR(1) representation captures the persistence in idiosyncratic volatilities and correlations. The assumption that all elements of the covariance matrix are driven by identical dynamics is obviously restrictive, however, yields parameter parsimony even in high dimensions. Estimation errors can therefore be strongly limited and parameter uncertainty can be straightforwardly captured by choosing appropriate prior distributios for μ_i, ϕ_i and σ_i. The approach can be seen as a strong parametric regularization of the covariance matrix which, however, still accommodates important empirical features. Furthermore, though the joint distribution of the data is conditionally Gaussian, the stationary distribution exhibits thicker tails.

The priors for the univariate stochastic volatility processes are chosen independently in line with Aguilar and West (2000) and Pati et al. (2014), i.e.,

$$p(\mu_i, \phi_i, \sigma_i) = p(\mu_i)p(\phi_i)p(\sigma_i). \tag{41}$$

The level μ_i is equipped with a normal prior, the persistence parameter ϕ_i is chosen such that $(\phi_i + 1)/2 \sim B(a_0, b_0)$, which enforces stationarity, and for σ_i^2 we assume $\sigma_i^2 \sim G\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$.[17] For each element of the factor loadings matrix, a hierarchical zero-mean Gaussian distribution is

[17] In the empirical application we set the prior hyper-parameters to $a_0 = 20, b_0 = 1.5$ and $B_\sigma = 1$ as proposed by Kastner et al. (2017).

chosen. Choosing hierarchical priors for the factor loadings allows us to enforce sparsity reducing parameter uncertainty especially in very high dimensions, see e.g. Griffin and Brown (2010) and Kastner (2016). Kastner (2016) proposing interweaving strategies in the spirit of Yu and Meng (2011) to reduce the enormous computational burden for high dimensional estimations of SV objects.

4.3 Covariance Shrinkage

The most simple and natural covariance estimator is the rolling window sample covariance estimator,

$$S_t := \frac{1}{h-1} \sum_{i=t-h}^{t} (r_i - \hat{\mu}_t)(r_i - \hat{\mu}_t)', \tag{42}$$

with $\hat{\mu}_t := \frac{1}{h-1} \sum_{i=t-h}^{t} r_i$, and estimation window of length h. It is well-known that S_t is highly inefficient and yields poor asset allocations as long as h does not sufficiently exceed N. To overcome this problem, Ledoit and Wolf (2004) propose shrinking S_t towards a more efficient (though biased) estimator of Σ_t.[18] The classical linear shrinkage estimator is given by

$$\hat{\Sigma}^{\text{t,Shrink}} = \hat{\delta} F_t + (1 - \hat{\delta}) S_t, \tag{43}$$

where F_t denotes the sample constant correlation matrix and $\hat{\delta}$ minimizes the Frobenius norm between F_t and S_t. For more details, see Ledoit and Wolf (2004). F_t is based on the sample correlations $\hat{\rho}_{ij} := \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$, where s_{ij} is the i-th element of the j-th column of the sample covariance matrix S_t. The average sample correlations are given by $\bar{\rho} := \frac{2}{(N-1)N} \sum_{i=1}^{N} \sum_{j=i+1}^{N-1} \hat{\rho}_{ij}$ yielding the ij-th element of F_t as $F_{t,ij} = \bar{\rho}\sqrt{\hat{\rho}_{ii}\hat{\rho}_{jj}}$.

Finally, the resulting predictive return distribution is obtained by assuming $p_t(r_{t+1}|\hat{\Sigma}^{t,Shrink}) \sim$

[18] Instead of shrinking the eigenvalues of S_t linearly, an alternative approach would be the implementation of non-parametric shrinkage in the spirit of Ledoit and Wolf (2012), which is left for future research.

$N(0, \hat{\Sigma}^{t,Shrink})$. Equivalently, a Gaussian framework is implemented for the sample variance-covariance matrix $p_t(r_{t+1}|S_t) \sim N(0, S_t)$. Hence, parameter uncertainty is only taken into account through the imposed regularization of the sample variance-covariance matrix. We refrain from imposing additional prior distributions to study the effect of a pure covariance regularization and to facilitate comparisons with the sample covariance matrix.

5 Empirical Analysis

5.1 Data and General Setup

In order to obtain a representative sample of US-stock market listed firms, we select all constituents from the S&P 500 index, which were traded during the complete time period starting in June 2007, the earliest date for which corresponding HF-data from the LOBSTER database is available. This results in a total dataset containing $N = 308$ stocks listed at Nasdaq.[19]

The data covers the period from June 2007 to March 2017, corresponding to 2,409 trading days after excluding weekends and holidays. Daily returns are computed based on end-of-day prices. All the computations are performed after adjusting for stock splits and dividends. As of March 2017, the total market capitalization of the stocks included in our setup exceeds 17,000 Billion USD, which comes close to the S&P 500 capitalization of 21,800 Billion USD. We extend our data set by HF data extracted from the LOBSTER database, which provides tick-level message data for every asset and trading day.[20] Utilizing midquotes amounts to more than 73 Billion observations.

[19]Exclusively focusing on stocks, which are continuously traded through the entire period, is a common proceeding in the literature and implies some survivorship bias and the negligence of younger companies with IPO's after 2007. In our allocation approach, this aspect could be in principle addressed by including *all* stocks from the outset and a priori imposing zero weights to stocks in periods, when they are not (yet) traded. Such a proceeding, however, would further increase the dimension of the asset space, implying additional computational burden. Moreover, the conclusions drawn from our empirical study are not critically dependent on this aspect. We therefore refrain from further extensions of the asset space.

[20]See https://lobsterdata.com.

Panel a) of Figure 3 visualizes the cross-sectional average of annualized realized volatilities estimated based on univariate realized kernels according to Barndorff-Nielsen et al. (2008) for each trading day. Panel b) of Figure 3 shows the average correlations computed using blocked realized kernel estimates based on 4 groups on a daily basis. We observe generally positive correlations revealing (partly) substantial and persistent fluctuations.

In order to investigate the prediction power and resulting portfolio performance of our models, we sequentially generate forecasts on a daily basis and compute the corresponding paths of portfolio weights. Table 1 summarizes the distinct steps of the estimation and forecasting procedure. We implement $K = 4$ different models as of Section 4. The HF approach is based on the BRK-Wishart model with 4 groups and the $t+1$ covariance matrix is predicted as the average of estimates over the most previous 5 days. The SV model is based on $j = 3$ common factors[21], while forecasts based on the sample variance-covariance matrix S_t and its shrunken version $\hat{\Sigma}^{\text{Shrink}}$ are based on a rolling window size of 500 days.

Every model is based on a Gaussian likelihood, where parameter uncertainty, as taken into account in the HF model and SV model, leads to fat-tailed predictive return distributions. Moreover, in line with a wide range of studies utilizing daily data, we refrain from predicting mean returns but assume $\mu_t = 0$ in order to avoid excessive estimation uncertainty. Therefore, we effectively perform global minimum variance optimization under transaction costs and parameter uncertainty as well as model uncertainty. In order to quantify the robustness and statistical significance of our results, we perform a bootstrap analysis by re-iterating the procedure described in Table 1 in total 200 times for random subsets of $N = 250$ stocks out of the total 308 assets.[22]

[21] The predictive accuracy is very similar for values of j between 1 and 5, but declines when including more factors.

[22] The forecasting and optimization procedures require substantial computing resources. As the portfolio weights at t depend on the allocation at day $t-1$, parallelization is restricted. Sequentially computing the multivariate realized kernels for every trading day, running the MCMC algorithm, performing numerical integration and optimizing in the high-dimensional asset space for all models requires approximately one month computing time on a strong cluster such as the Vienna Scientific Cluster.

Bootstrap iteration	Randomly choose investment horizon $N = 250$ out of all 308 assets
Initial	Choose predictive models $\mathcal{D} := (\mathcal{M}_1, \ldots, \mathcal{M}_K)$ Initialize weights $\hat{\omega}_0^{\mathcal{M}_k} := \frac{1}{N}\iota \; \forall k = 1, \ldots, 4$
At $t > 0$	For every model $\mathcal{M}_k$ 1. MCMC-sampling from posterior distribution $\pi(\Theta_k \| R_t, \mathcal{H}_t, \mathcal{M}_k)$ 2. Generate $L = 100.000$ draws from predictive $p_t(r_{t+1} \| \mathcal{M}_k)$ 3. Compute optimal $\hat{\omega}_{t+1}^{\mathcal{M}_k}$ depending on current allocation $\omega_{t^+}^{\mathcal{M}_k}$ 4. Evaluate (predictive) performance $p(r_t \| R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k)$
	For the mixture model 1. Update weight c_t and generate samples from $p(r_{t+1} \| R_t, \mathcal{H}_t, \mathcal{D})$ 2. Compute allocation vector $\hat{\omega}_{t+1}^{\mathcal{D}}$ (using Equation (EU))
At t = T	Evaluate the portfolio performances of bootstrap iteration
Bootstrap iteration	Go back to step 1 (and repeat this 200 times)

Table 1: Schematic illustration of the procedure to compute portfolio weights taking into account parameter uncertainty, transaction costs and model mixing in our framework.

5.2 Evaluation of the Predictive Performance

In a first step, we evaluate the predictive performance of the individual models. This does not require computing portfolio weights and thus is based on steps 1 and 2 according to Table 1. In order to visualize the model's forecasts of the high-dimensional return distribution, we use the generated samples $r_{t+1}^{\mathcal{M}_k,(l)}$ from the posterior predictive distribution $p_t(r_{t+1}|\mathcal{M}_k)$ and compute samples of the predicted return distribution of the naive portfolio $\omega_{\text{naive}}' r_{t+1}^{\mathcal{M}_k,(l)}$. Figure 4 visualizes the 95% credibility regions implied by the simulated predictive distribution of the day $t+1$ (naive) portfolio return. The dots indicate observed returns of the naive portfolio at day $t+1$, whereas the lines indicate the 95% credible regions of the individual models.

We observe substantial differences between the models. The SV model reveals the highest fluctuations, reflecting the noisiness in daily returns. On the other hand, the flexibility of the SV model to capture idiosyncratic changes in volatilities implies fast responses to changing market conditions. During the financial crisis, for instance, the SV model implies a substantial widening of the credible regions, indicating high market volatility. In contrast, predictions stemming from the HF model are less volatile. This is partly due to the 5-day smoothing which reduces the noisiness of forecasts, as it scales down the impact of single days of over-shooting volatility.[23] Nevertheless, these forecasts are responsive and obviously quickly react to changing market conditions. Finally, forecasts implied by the shrinkage estimator change only slowly through time. This is not surprising as this estimator builds on rolling two-year windows of daily data and thus updates only very slowly to new information.

	Mean	SD	Best
HF	707.0	117.08	0.171
	[702.9,711.0]	[113.58,119.81]	[0.160,0.181]
Sample	586.5	337.56	0.005
	[580.7,591.8]	[318.99,353.23]	[0.003,0.007]
LW	694.2	165.72	0.197
	[689.1,699.1]	[161.29,170.00]	[0.185,0.211]
SV	705.0	138.30	0.135
	[701.6,708.2]	[136.11,140.10]	[0.126,0.146]
Combination	731.3	97.95	0.484
	[727.1,735.6]	[96.71,99.34]	[0.477,0.491]

Table 2: Predictive accuracy of the models, evaluated using the time series of out-of-sample predictive log likelihoods, corresponding to $\log p_t(r_{t+1}|\mathcal{M}_k)$. The values in brackets correspond to bootstrapped 95 % confidence intervals. *Mean* denotes the average posterior predictive log-likelihood. *SD* is the standard deviation of the time series and *Best* is the fraction of total days where the individual model obtained the highest predictive accuracy among its competitors. *Combination* corresponds to predictions created by combining the individual models based on (24).

Evaluating the predictive performance of high-dimensional return distributions is not straightforward. A popular metric is the log posterior predictive likelihood $\log p(r_t^O|R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k)$, where r_{t+1}^O are the observed returns at day t, indicating how much probability mass the predictive distribution assigns to the observed outcomes, see e.g., Weigend and Shi (2000), Amisano

[23] Without smoothing, HF-data based forecasts would be prone to substantial higher fluctuations, especially on days with extraordinary intra-daily activity such as on the Flash Crash in May 2010. We find that these effects reduce the predictive ability.

and Giacomini (2007) and Bao et al. (2007). Table 2 gives summary statistics of the (daily) time series of the *out-of-sample* log posterior predictive likelihood $\log p(r_t^O|R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k)$ for each model.

In terms of the mean posterior predictive log-likelihood obtained in our sample, the sample variance-covariance matrix solely is not sufficient to provide accurate forecasts. Shrinking the covariance matrix, however, significantly increases its forecasting performance. Both estimators, however, still significantly under-perform the SV and HF model. The fact that the SV model performs widely similar to the HF model is an interesting finding as it utilizes only daily data and thus much less information than the HF model. This disadvantage, however, seems to be overcompensated by the fact that the model captures daily dynamics in the data and thus straightforwardly produces one-step-ahead forecasts. In contrast, the HF model produces accurate estimates of Σ_t, but does not allow for any projections into the future. Our results show that both the efficiency of estimates Σ_t and the incorporation of daily dynamics are obviously crucial for superior out-of-sample predictions. The fact that both the SV model and the HF model perform widely similar indicates that the respective advantages and disadvantages of the individual models counterbalance each other. We thus expect that an appropriate dynamic forecasting model for vast-dimensional covariances, e.g., in the spirit of Hansen et al. (2012) may perform even better in terms of the mean posterior predictive log-likelihood, but may contrariwise induce additional parameter uncertainty.[24]

The last row of Table 2 gives the obtained out-of-sample predictive performance of the model combination approach as discussed in Section 3. Computing combination weights c_t^* as described in Equation (24) and evaluating the predictive density computing

$$LS_{t+1}^{\text{Combination}} = \log \left(\sum_{k=1}^{K} c_{t,k} p \left(r_{t+1}^O | R_t, \mathcal{H}_t, \mathcal{M}_k \right) \right)$$

[24]Given that it is not straightforward to implement such a model in the given general and high-dimensional setting, we leave such an analysis to future research.

reflects a significantly stronger prediction performance through the entire sample. Combining both high- and low-frequency based approaches thus increases the predictive accuracy and outperforms all individual models.

Figure 5 depicts the time series of resulting model combination weights, reflecting the relative past prediction performance of each model at each day. We observe distinct time variations, which are obviously driven by the market environment. The gray shaded area in Figure 5 shows the daily averages of the estimated assets' volatility, computed using univariate realized kernels. We thus observe that during high-volatility periods, the HF approach produces superior forecasts and has the highest weight. This is particularly true during the financial crisis and during more recent periods of market turmoil, where the estimation precision induced by HF data clearly pays off. Conversely, forecasts based on daily stochastic volatility perform considerably strong in more calm periods as in 2013/2014. Forecasts implied by the shrinkage estimator have lower but non-negligible weight, and thus significantly contribute to an optimal forecast combination. In contrast, predictions based on the sample variance-covariance matrix are negligible and are always dominated by the shrinkage estimator.

Superior predictive accuracy, however, is not equivalent to superior portfolio performance. The volatility of the predictive credible regions visualized in Figure 4 underlines that an investor *not* adjusting for transaction costs ex ante may re-balance her portfolio unnecessarily often if she relies on the predictions based on the stochastic volatility factor model and – to less extent – on the HF-based forecasts. Though the underlying predictions may be quite accurate, transaction costs can easily offset this advantage compared to the use of the rather smooth predictions implied by (regularized) forecasts based on rolling windows. This aspect will analyzed in the following section.

5.3 Evaluation of Portfolio Performance

To evaluate the performance of the resulting portfolios implied by the individual models' forecasts, we follow the procedure described in Table 1 and construct portfolios based on bootstrapped portfolio weights. Accordingly, the underlying asset universe consists of 250 assets which are randomly drawn out of the entire asset space.

Our setup represents an investor using the available information to sequentially update her beliefs about the parameters and state variables of the return distribution of the 250 selected assets. Based on the estimates, she generates predictions of the returns of tomorrow and accordingly allocates her wealth by solving (EU), using risk aversion $\gamma = 4$. After holding the assets for a day, she realizes the gains and losses, updates the posterior distribution and re-computes optimal portfolio weights. This procedure is repeated for each period and allows analyzing the time series of the realized ("out-of-sample") returns $r^k_{t+1} = \sum_{i=1}^{N} \hat{\omega}^{\mathcal{M}_k}_{t+1,i} r^O_{t+1,i}$. Bootstrapping allows us to investigate the realized portfolio performances of 200 different investors, differing only with respect to the available set of assets.

We assume proportional (L_1) transaction costs according to Equation (9). We choose this parametrization, as it is a popular choice in the literature, see, e.g., DeMiguel et al. (2009), and is more realistic than quadratic transaction costs as studied in Section 2. As suggested by DeMiguel et al. (2009), we fix β to 50bp, corresponding to a rather conservative proxy for transaction costs on the U.S. market. Though such a choice is only a rough approximation to real transaction costs, which in practice depend on (possibly time-varying) institutional rules and the liquidity supply in the market, we do not expect that our general findings are specifically driven by this choice. While it is unavoidable that transaction costs are underestimated or overestimated in individual cases, we expect that the overall effects of turnover penalization can be still captured with realistic magnitudes. Our results in Chapter 2 moreover reflect to which extent the portfolio performance changes in dependence of β and reveals a certain robustness of the effects for values of β around 50bp.

The returns resulting from evaluating realized performances net of transaction costs are then given by

$$r_t^{k,\text{nTC}} = r_t^k - \nu_t\left(\hat{\omega}_{t+1}^{\mathcal{M}_k}\right) = \sum_{i=1}^{N} \hat{\omega}_{t,i}^{\mathcal{M}_k} r_{t,i}^{O} - \nu_t\left(\hat{\omega}_t^{\mathcal{M}_k}\right), \quad t = 1, \ldots, T. \tag{44}$$

We quantify the portfolio performance based on the average portfolio return, its volatility, its Sharpe ratio and the certainty equivalent for $\gamma = 4$:

$$\hat{\mu}\left(r^{k,\text{nTC}}\right) := \frac{1}{T}\sum_{t=1}^{T} r_t^{k,\text{nTC}}, \tag{45}$$

$$\hat{\sigma}\left(r^{k,\text{nTC}}\right) := \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(r_t^{k,\text{nTC}} - \hat{\mu}\left(r^{k,\text{nTC}}\right)^2\right)}, \tag{46}$$

$$SR^k := \frac{\hat{\mu}\left(r^{k,\text{nTC}}\right)}{\hat{\sigma}\left(r^{k,\text{nTC}}\right)}, \tag{47}$$

$$CE^k := 100\left(\frac{1}{T}\sum_{t=1}^{T}\left(r_t^{k,\text{nTC}}\right)^{1-4} - 1\right)^{\frac{1}{1-4}}. \tag{48}$$

Moreover, we quantify the portfolio turnover

$$TO^k := \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|\hat{\omega}_{t,i}^{\mathcal{M}_k} - \frac{\hat{\omega}_{t-1,i}^{\mathcal{M}_k} \circ r_{t,i}}{\iota'(\hat{\omega}_{t-1}^{\mathcal{M}_k} \circ r_t)}\right\|_1, \tag{49}$$

the average weight concentration,

$$pc^k := \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\hat{\omega}_{t,i}^{\mathcal{M}_k}\right)^2, \tag{50}$$

and the average size of the short positions as given by

$$sp^k := \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}|\hat{\omega}_{t,i}|\mathbb{1}_{\left\{\hat{\omega}_{t,i}^{\mathcal{M}_k}<0\right\}}. \tag{51}$$

Finally, we compare the performance of the resulting (optimal) portfolios to those of a number of benchmark portfolios. First, we implement the naive portfolio allocation $\omega^{\text{Naive}} := \frac{1}{N}\iota$ based on daily re-balancing.[25]

Moreover, we include the "classical" global minimum variance portfolio based on the Ledoit-Wolf shrinkage estimator, $\omega_{t+1}^{\text{mvp}} := \frac{\left(\hat{\Sigma}^{\text{t, Shrink}}\right)^{-1}\iota}{\iota'\left(\hat{\Sigma}^{\text{t, Shrink}}\right)^{-1}\iota}$. Furthermore, as proposed by Jagannathan and Ma (2003), we compute optimal global minimum variance weights with a no-short sale constraint, $\omega_{t+1}^{\text{mvp, no s.}}$, computed as solution to the optimization problem:

$$\omega_{t+1}^{\text{mvp, s}} = \arg\min \omega'\hat{\Sigma}^{\text{t, Shrink}}\omega \text{ s.t. } \iota'\omega = 1 \text{ and } \omega_i \geq 0 \;\forall i = 1,\ldots,N.$$

We also include the Bayes-Stein estimator as proposed by Jorion (1986), for which we do not report results as the approach is not able to outperform the naive portfolio in our sample.

Table 3 summarizes the results. The results in the first panel correspond to strategies ignoring transaction costs in the portfolio optimization by setting $v_t(\omega) = 0$ when computing optimal weights ω^*. These strategies employ the information conveyed by the predictive models, but ignore transaction costs. However, after re-balancing, the resulting portfolio returns are computed net of transaction costs. This corresponds to common proceeding, where turnover penalization is done ex post, but is not incorporated in the decision process.

As indicated by highly negative average portfolio returns, a priori turnover penalizing is crucial in order to obtain reasonable portfolio performances in a setup with transaction costs. The major reason is not necessarily the use of a sub-optimal weight vector, but rather the

[25] We also implemented a weekly and monthly re-balancing, which, however, does not qualitatively alter the results.

	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	TO	pc	sp	% Trade
$\omega^{\text{HF, no TC}}$	-96.5	15.90	-	-	613.49	0.13	0.08	0.8835
$\omega^{\text{SV, no TC}}$	-92.3	13.77	-	-	401.63	0.54	0.41	0.9362
$\omega^{\text{LW, no TC}}$	-58.7	16.09	-	-	139.51	0.35	0.18	0.8499
$\omega^{\text{Sample, no TC}}$	-48.4	15.92	-	-	113.10	0.59	1.50	0.8548
ω^{HF}	6.4	15.77	0.407	6.274	1.70	0.49	1.11	0.0804
ω^{SV}	6.7	16.74	0.400	6.271	0.25	0.44	0.94	0.0367
ω^{LW}	6.6	16.67	0.396	6.251	0.24	0.24	0.13	0.0346
ω^{Sample}	5.5	16.47	0.335	5.170	0.47	0.23	0.13	0.0367
$\omega^{\text{Combination}}$	6.6	14.83	0.443	6.356	1.22	0.20	0.15	0.1208
ω^{Naive}	5.0	23.48	0.215	4.759	2.76	0.45	0.00	0.9184
ω^{mvp}	-59.0	14.10	-	-	133.56	0.90	1.21	0.9996
$\omega^{\text{mvp no s.}}$	2.7	13.09	0.209	2.468	8.82	0.01	0.00	0.9996

Table 3: Annualized averages of the bootstrapped out-of-sample portfolio performances based on 1904 trading days. Transaction costs are proportional to the L_1 norm of re-balancing (as of (9)). $\hat{\mu}$ is the annualized portfolio return in percent, $\hat{\sigma}$ the annualized standard deviation in percent, SR denotes the (annualized) out-of-sample Sharpe ratio of the individual strategies, CE is the Certainty Equivalent for an investor with power utility function and risk-aversion factor $\gamma = 4$, TO is the average turnover in percent, pc is the average portfolio concentration (L_2 norm of the portfolio weights) and sp is the average proportion of the sum of negative portfolio weights. *% Trade* is the fraction of days with trading activity more than 0.001%.

fact that these portfolio positions suffer from extreme turnover due to a high responsiveness of the underlying covariance predictions to changing market conditions, thus implying frequent re-balancing. All four predictive models generate average annualized portfolio turnover of more than 100%, which sums up to substantial losses during the trading period. If an investor would have started trading with 100 USD in June 2007 using the HF-based forecasts without adjusting for transaction costs, she would end up with less than 0.01 USD in early 2017. The worse performance of HF-based and SV-based predictions compared to the sample covariance and shrinkage estimators could already have been anticipated by the strong fluctuation of the predictions illustrated in Figure 4 as they produce a significantly higher turnover. Moreover, none of the four approaches is able to outperform the naive portfolio, though the individual predictive models clearly convey information. We conclude that the adverse effect of high turnover becomes particularly strong in case of large-dimensional portfolios.[26]

Explicitly adjusting for transaction costs, however, strongly changes the picture: The implemented strategies produce significantly positive average portfolio returns and reasonable Sharpe ratios. The turnover is strongly reduced and amounts to less than 1% of the turnover implied by strategies which do not account for transaction costs. All strategies clearly outperform the naive and GMV strategies in terms of net-of-transaction-cost Sharpe ratios and certainty equivalents.

Comparing the performance of individual models, the HF-based model, the SV model and the shrinkage estimator yield the strongest portfolio performances, but perform very similarly in terms of SR and CE. In line with the theoretical findings in Section 2, we thus observe that turnover regularization reduces the performance differences between individual models. Nevertheless, *combining forecasts*, however, outperforms individual models and yields Sharpe ratios which are approximately 10% higher. We conclude that the combination of predictive distributions resulting from HF data with those resulting from low-frequency (i.e., daily) data is beneficial – even after accounting for transaction costs. Not surprisingly, the sample covariance performs worse but still provides a reasonable Sharpe ratio. This confirms the findings in Section 2 and illustrates that under turnover regularization even the sample covariance yields sensible inputs for high-dimensional predictive return distributions.

Imposing restrictions to reduce the effect of estimation uncertainty, such as a no-short sale constraint in GMV optimization ($\omega^{\text{mvp no s.}}$), however, does not yield a competing performance.

[26] Bollerslev et al. (2016) find reverse results for GMV portfolio optimization based on HF-based covariance forecasts. For $N = 10$ assets, they find an over-performance of HF-based predictions even in the presence of transaction costs. Two reasons may explain the different findings: First, the burden of a high dimensionality implies very different challenges when working with more than 300 assets. In addition, Bollerslev et al. (2016) employ methods, which directly impose a certain regularization. In light of our findings in Section 2, this can be interpreted as making the portfolios (partly) robust to transaction costs even if the latter are not explicitly taken into account.

These findings underline our conclusions drawn from Proposition 3: Though gross-exposure constraints are closely related to a penalization of transaction costs and minimize the effect of misspecified elements in the variance-covariance matrix, see, e.g., Fan et al. (2012), such a regularization yields sub-optimal weights in the actual presence of transaction costs.

The last column of Table 3 gives the average fraction of days where the portfolio is re-balanced.[27] We observe that the incorporation of transaction costs reduces this percentage from approximately 90% to less than 5%. In case of shrinkage-based forecast, for instance, the portfolio is re-balanced only roughly 9 times a year. Figure 6 shows the time series of daily turnover activity (TO^k, measured as the L_1 norm of re-balancing and therefore directly related to daily transaction costs) for the individual strategies. The HF strategy, for instance, trades frequently during the financial crisis, but less than 10 times since 2012. Turnover penalization can be thus interpreted as a Lasso-type mechanism, which enforces buy-and-hold strategies over longer periods as long as markets do not reveal substantial shifts requiring re-balancing. In contrast, the benchmark strategies that do not account for transaction costs tend to re-balance permanently (small fractions of) wealth, thus cumulating an extraordinarily high amount of transaction costs. For instance, as shown by Figure 6, the naive portfolio re-balances constantly, causing an average turnover which is higher than for all of the penalized strategies.

5.3.1 Economic significance of the portfolio performance

To interpret the economic significance of out-of-sample portfolio performances, we evaluate the resulting utility gains using the framework by Fleming et al. (2003). We thus compute the hypothetical fees, an investor with power utility and a relative risk aversion $\gamma = 4$ would be willing to pay on an annual basis to switch from an individual strategy $\mathcal{M}_1$ to strategy $\mathcal{M}_2$.[28] The fee is computed such that the investor would be indifferent between the two strategies in

[27] Due to numerical instabilities, we interpret a turnover of less than 0.001% as no re-balancing.
[28] We also used alternative values of γ with $\gamma = 1, \ldots, 10$, however, do not find qualitative differences.

terms of the resulting utility. For $r_t^{\mathcal{M}_1,nTC}$ and $r_t^{\mathcal{M}_2,nTC}$, we thus determine $\Delta_{\mathcal{M}_1,\mathcal{M}_2}$ such that

$$\sum_{t=1}^{T}(r_t^{\mathcal{M}_1,nTC})^{(1-\gamma)} = \sum_{t=1}^{T}(r_t^{\mathcal{M}_2,nTC} - \Delta_{\mathcal{M}_1,\mathcal{M}_2})^{(1-\gamma)}. \tag{52}$$

Figure 7 shows the bootstrap-based distributions of the resulting fees to switch from the individual strategies to the model combination strategy. We thus find that on average, investors would be willing to pay positive amounts to switch to the superior mixing models. For none of the 6 different strategies, the 5% quantiles of the performance fees are below zero, indicating the strong performance of the mixing model. Hence, even after transaction costs investors would gain higher utility by combining forecasts based on high- and low-frequency data.

In our high-dimensional setup, an investor would be willing to pay 7.5 basis points on average on an annual basis to switch from the naive allocation to the leading model combination approach. The $1/N$ allocation is thus not only statistically but also economically inferior. However, as expected based on the findings in Section 2, switching fees between the individual penalized models are substantially lower. Table 4 shows the average fees an investor would be willing to pay to switch from strategy $\mathcal{M}_i$ in column i to strategy $\mathcal{M}_j$ in row j. Hence, we observe that the marginal gain of switching from the HF-based approach to the SV model amounts to just 0.1 bp per year. This result again confirms the finding that relative performance differences between high-frequency-based and low-frequency-based approaches level out when transaction costs are accounted for.

6 Conclusions

This paper theoretically and empirically studies the role of transaction costs in large-scale portfolio optimization problems. We show that the ex ante incorporation of transaction costs regularizes the underlying covariance matrix. The implied turnover penalization improves portfolio

	Naive	HF	Sample	LW	SV	Combination
mvp no s.	- 1.9	2.5	2.1	2.3	2.0	2.7
Naive		6.4	5.9	6.1	5.8	7.5
HF			-0.3	0.1	0.1	1.1
Sample				0.0	0.1	1.3
LW					0.4	1.4
SV						2.4

Table 4: Average annual performance fees in basis points for switching between the individual models based on $\gamma = 4$. The table reads: *On an annual basis an investor would be willing to pay 7.5 basis point to switch from the naive portfolio to the combination strategy.*

allocations in terms of Sharpe ratios and utility gains. This is on the one hand due to regularization effects improving the stability and conditioning of covariance matrix estimates and on the other hand due to a significant reduction of the amount (and frequency) of re-balancing and thus turnover costs.

In contrast to a pure (statistical) regularization of the covariance matrix, in case of turnover penalization, the regularization parameter is naturally given by the transaction costs prevailing in the market. This a priori institution-implied regularization reduces the need for exclusive covariance regularizations (as, e.g., implied by shrinkage), but does not make it superfluous. The reason is that a transaction cost regularization only partly contributes to a better conditioning of covariance estimates, but does not guarantee this effect at first place. Accordingly, procedures which additionally stabilize predictions of the covariance matrix and their inverse, are still effective but are less crucial as in the case where transaction costs are neglected.

Performing an extensive study utilizing Nasdaq data of more than 300 assets from 2007 to 2017, we empirically show the following results: First, we find that neither high-frequency-based nor low-frequency-based predictive distributions result into positive Sharpe ratios when transaction costs are *not* taken into account ex ante. This is mainly due to a high turnover implied by (too) frequent re-balancing. Second, as soon as transaction costs are incorporated ex ante into the optimization problem, resulting portfolio performances strongly improve, but differences in the relative performance of competing predictive models become smaller. In particular, we find that predictions implied by HF-based (blocked) realized covariance kernels, by a daily factor SV model and by a rolling window shrinkage approach perform statistically and economically widely similarly. A portfolio bootstrap reveals that none of these approaches is able to produce significant utility gains on top of one of the competing models. Third, despite of a similar performance of individual predictive models, mixing HF and low-frequency information is beneficial as it exploits the time-varying nature of the individual model's predictive ability. We find that HF-based predictions are superior in volatile market periods, but are dominated by SV-based predictions utilizing daily data in calm periods. Finally, we show that in case of an ex-ante consideration of transaction costs, naive ($1/N$) strategies or GMV strategies (without turnover penalization) are significantly outperformed. We moreover show that restrictions on the portfolio weights as proposed by Jagannathan and Ma (2003) do not yield competing strategies as they do not imply (sufficient) turnover penalization.

Our paper thus shows that transaction costs play a substantial role for portfolio allocation and reduce the benefits of individual predictive models. Nevertheless, it pays off to optimally combine high-frequency and low-frequency information. Allocations generated by adaptive mixtures dominate strategies from individual models as well as any naive strategies.

References

Acharya, V. V. and Pedersen, L. H. (2005), 'Asset pricing with liquidity risk', *Journal of Financial Economics* **77**(2), 375–410.

Aguilar, O. and West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business & Economic Statistics* **18**(3), 338–357.

Amisano, G. and Giacomini, R. (2007), 'Comparing Density Forecasts via Weighted Likelihood Ratio Tests', *Journal of Business & Economic Statistics* **25**(2), 177–190.

Andersen, T. G. and Bollerslev, T. (1998), 'Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts', *International Economic Review* **39**(4), 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **96**(453), 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Anderson, E. W. and Cheng, A.-R. (2016), 'Robust bayesian portfolio choices', *Review of Financial Studies* **29**(5), 1330–1375.

Andersson, J. (2001), 'On the normal inverse Gaussian stochastic volatility model', *Journal of Business & Economic Statistics* **19**(1), 44–54.

Avramov, D. (2003), 'Stock return predictability and asset pricing models', *Review of Financial Studies* **17**(3), 699–738.

Avramov, D. and Zhou, G. (2010), 'Bayesian portfolio analysis', *Annual review of financial economics* **2 (2010)**, 25–47.

Bao, Y., Lee, T.-H. and Saltoğlu, B. (2007), 'Comparing density forecast models', *Journal of Forecasting* **26**(3), 203–225.

Barndorff-Nielsen, O. E. (1997), 'Normal inverse Gaussian distributions and stochastic volatility modelling', *Scandinavian Journal of Statistics* **24**(1), 1–13.

Barndorff-Nielsen, O. E. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society: Series B* **64**(2), 253–280.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* **12**(3), C1–C32.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011), 'Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. E., Hansen, P. R., Shephard, N. and Lunde, A. (2008), 'Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise', *Econometrica* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2004), 'Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics', *Econometrica* **72**(3), 885–925.

Bates, J. M. and Granger, C. W. J. (1969), 'The combination of forecasts', *OR* pp. 451–468.

Billio, M., Casarin, R., Ravazzolo, F. and van Dijk, H. K. (2013), 'Time-varying combinations of predictive densities using nonlinear filtering', *Journal of Econometrics* **177**(2), 213–232.

Bollerslev, T., Patton, A. J. and Quaedvlieg, R. (2016), 'Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions', *Working Paper* .

Brandt, M. W., Santa-Clara, P. and Valkanov, R. (2009), 'Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns', *Review of Financial Studies* **22**(9), 3411–3447.

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

Brown, S. J. (1976), Optimal portfolio choice under uncertainty: a Bayesian approach, PhD Thesis, University of Chicago, Chicago.

Chib, S. and Greenberg, E. (1995), 'Understanding the metropolis-hastings algorithm', *The American Statistician* **49**(4), 327–335.

Chib, S., Nardari, F. and Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**(2), 341–371.

Chopra, V. K. and Ziemba, W. T. (1993), 'The effect of errors in means, variances, and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**(2), 6–11.

Davis, M. H. A. and Norman, A. R. (1990), 'Portfolio selection with transaction costs', *Mathematics of Operations Research* **15**(4), 676–713.

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

Durham, G. and Geweke, J. (2014), 'Improving asset price prediction when all models are false', *Journal of Financial Econometrics* **12**(2), 278–306.

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

Engle, R. F., Ledoit, O. and Wolf, M. (2017), 'Large dynamic covariance matrices', *Journal of Business & Economic Statistics* (Forthcoming).

Fan, J., Fan, Y. and Lv, J. (2008), 'High dimensional covariance matrix estimation using a factor model', *Journal of Econometrics* **147**(1), 186–197.

Fan, J., Zhang, J. and Yu, K. (2012), 'Vast Portfolio Selection with Gross-exposure Constraints', *Journal of the American Statistical Association* **107**(498), 592–606.

Fleming, J., Kirby, C. and Ostdiek, B. (2003), 'The economic value of volatility timing using "realized" volatility', *Journal of Financial Economics* **67**(3), 473–509.

Forsberg, L. and Bollerslev, T. (2002), 'Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): The GARCH-NIG model', *Journal of Applied Econometrics* **17**(5), 535–548.

Garlappi, L., Uppal, R. and Wang, T. (2007), 'Portfolio selection with parameter and model uncertainty: A multi-prior approach', *Review of Financial Studies* **20**(1), 41–81.

Gârleanu, N. (2009), 'Portfolio choice and pricing in illiquid markets', *Journal of Economic Theory* **144**(2), 532–564.

Gârleanu, N. and Pedersen, L. H. (2013), 'Dynamic trading with predictable returns and transaction costs', *Journal of Finance* **68**(6), 2309–2340.

Geweke, J., Koop, G. and Dijk, Herman K. van, eds (2011), *The Oxford Handbook of Bayesian Econometrics*, Oxford Handbooks in Economics, Oxford Univ. Press, Oxford and New York.

Griffin, J. E. and Brown, P. J. (2010), 'Inference with normal-gamma prior distributions in regression problems', *Bayesian Analysis* **5**(1), 171–188.

Han, Y. (2006), 'Asset allocation with a high dimensional latent factor stochastic volatility model', *Review of Financial Studies* **19**(1), 237–271.

Hansen, P. R., Huang, Z. and Shek, H. H. (2012), 'Realized garch: A joint model for returns and realized measures of volatility', *Journal of Applied Econometrics* **27**(6), 877–906.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.

Hautsch, N., Kyj, L. M. and Malec, P. (2015), 'Do High–Frequency Data Improve High–Dimensional Portfolio Allocations?', *Journal of Applied Econometrics* **30**(2), 263–290.

Hautsch, N., Kyj, L. M. and Oomen, R. C. A. (2012), 'A blocking and regularization approach to high–dimensional realized covariance estimation', *Journal of Applied Econometrics* **27**(4), 625–645.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian Model Averaging: A Tutorial', *Statistical Science* **14**(4), 382–401.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2002), 'Bayesian analysis of stochastic volatility models', *Journal of Business & Economic Statistics* **20**(1), 69–87.

Jagannathan, R. and Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1684.

Jensen, M. B. and Lunde, A. (2001), 'The NIG–S&ARCH model: A fat–tailed, stochastic, and autoregressive conditional heteroskedastic volatility model', *Econometrics Journal* **4**(2), 319–342.

Jin, X. and Maheu, J. M. (2013), 'Modeling Realized Covariances and Returns', *Journal of Financial Econometrics* **11**(2), 335–369.

Jobson, J. D., Korkie, B. and Vinod, R. (1979), 'Improved estimation for Markowitz portfolios using James-Stein type estimators', *Proceedings of the Business and Economic Statistics Section* pp. 279–284.

Johannes, M., Korteweg, A. and Polson, N. (2014), 'Sequential learning, predictability, and optimal portfolio returns', *Journal of Finance* **69**(2), 611–644.

Jorion, P. (1986), 'Bayes-Stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**(03), 279–292.

Kan, R. and Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(03), 621–656.

Kastner, G. (2016), 'Sparse Bayesian time-varying covariance estimation in many dimensions', *ArXiv e-prints* .

Kastner, G., Frühwirth-Schnatter, S. and Lopes, H. F. (2017), 'Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models', *Journal of Computational and Graphical Statistics* (Forthcoming).

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**(5), 603–621.

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

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

Liu, Q. (2009), 'On portfolio optimization: How and when do we benefit from high–frequency data?', *Journal of Applied Econometrics* **24**(4), 560–582.

Lunde, A., Shephard, N. and Sheppard, K. (2016), 'Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice', *Journal of Business & Economic Statistics* **34**(4), 504–518.

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

Pati, D., Bhattacharya, A., Pillai, N. S. and Dunson, D. (2014), 'Posterior contraction in sparse Bayesian factor models for massive covariance matrices', *Annals of Statistics* **42**(3), 1102–1130.

Pettenuzzo, D. and Ravazzolo, F. (2016), 'Optimal Portfolio Choice Under Decision-Based Model Combinations', *Journal of Applied Econometrics* p. forthcoming.

Pflug, G. C., Pichler, A. and Wozabal, D. (2012), 'The 1/N investment strategy is optimal under high model ambiguity', *Journal of Banking & Finance* **36**(2), 410–417.

Shephard, N. (1996), 'Statistical aspects of ARCH and stochastic volatility', *Monographs on Statistics and Applied Probability* **65**, 1–68.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.

Tu, J. and Zhou, G. (2010), 'Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty', *Journal of Financial and Quantitative Analysis* **45**(04), 959–986.

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.

Uppal, R. and Wang, T. (2003), 'Model Misspecification and Underdiversification', *Journal of Finance* **58**(6), 2465–2486.

Wang, Z. (2005), 'A shrinkage approach to model uncertainty and asset allocation', *Review of Financial Studies* **18**(2), 673–705.

Weigend, A. S. and Shi, S. (2000), ‘Predicting daily probability distributions of S&P500 returns’, *Journal of Forecasting* **19**(4), 375–392.

Yu, Y. and Meng, X.-L. (2011), ‘To Center or Not to Center: That Is Not the Question—An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency’, *Journal of Computational and Graphical Statistics* **20**(3), 531–570.

A Proofs and Derivations

Proof of Proposition 1. First, we proof the following Lemma.

Lemma 1. *For $\beta \to \infty$ it holds that*

$$\lim_{\beta\to\infty} \beta {\Sigma^*}^{-1} = \gamma I \tag{53}$$

Proof. Note that any $N \times N$ matrix Σ cannot have eigenvalues of arbitrary magnitude, so $\Sigma^* := \left(\frac{\beta}{\gamma} I + \Sigma\right)^{-1}$ is defined for all β in some neighborhood of infinity. We have

$$\beta \left(\frac{\beta}{\gamma} I + \Sigma\right)^{-1} = \beta \left(\beta \left(\frac{1}{\gamma} I + \frac{1}{\beta}\Sigma\right)\right)^{-1} = \left(\frac{1}{\gamma} I + \frac{1}{\beta}\Sigma\right)^{-1}. \tag{54}$$

If we define $f(h) = \left(\frac{1}{\gamma} I + h\Sigma\right)^{-1}$, we are interested in the limit $\lim_{h\to 0} f(h)$. $f(h)$ is defined in a neighborhood around 0 and since taking inverses is continuous, we have

$$\lim_{\beta\to\infty} \beta {\Sigma^*}^{-1} = \lim_{h\to 0} f(h) = f(0) = \left(\frac{1}{\gamma} I + 0 \cdot \Sigma\right)^{-1} = \gamma I. \tag{55}$$

□

Using the classical mean-variance efficient portfolio representation

$$\omega^* = \frac{1}{\gamma}\left(\Sigma^{*-1} - \frac{1}{\iota'\Sigma^{*-1}\iota}\Sigma^{*-1}\iota\iota'\Sigma^{*-1}\right)\mu^* + \frac{1}{\iota'\Sigma^{*-1}\iota}\Sigma^{*-1}\iota \tag{56}$$

$$= \frac{1}{\gamma}\left(\beta\Sigma^{*-1} - \frac{1}{\iota'\beta\Sigma^{*-1}\iota}\beta\Sigma^{*-1}\iota\iota'\beta\Sigma^{*-1}\right)\underbrace{\frac{1}{\beta}\mu^*}_{=\frac{1}{\beta}\mu+\omega_{t^+}} + \frac{1}{\iota'\Sigma^{*-1}\iota}\Sigma^{*-1}\iota \tag{57}$$

we can derive the following equation based on Lemma 1:

$$\lim_{\beta\to\infty}\omega^* = \left(I - \frac{1}{N}\iota\iota'\right)\omega_{t^+} + \frac{1}{N}\iota \tag{58}$$

$$= \omega_{t^+}. \tag{59}$$

For increasing transaction costs optimal re-balancing shrinks to 0. □

Proof of Proposition 2. First, we can rewrite $\left(\frac{\beta}{\gamma}A(\Sigma^*)\right) = \left(\tilde{\Sigma}^{-1} - \frac{1}{\iota'\tilde{\Sigma}^{-1}\iota}\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}\right)$ with $\tilde{\Sigma} := \left(I + \frac{\gamma}{\beta}\Sigma\right)$. Let $\lambda_1 > \ldots > \lambda_N > 0$ the N eigenvalues of Σ. Then $(1 + \frac{\gamma}{\beta}\lambda_i), i = 1, \ldots, N$ are the eigenvalues of $\tilde{\Sigma}$ and the eigenvalues of $\tilde{\Sigma}^{-1}$ are $\frac{\beta}{\gamma}\frac{1}{\frac{\beta}{\gamma}+\lambda_i}$. Next we derive

$$\left\|\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)\right\|_F = \left\|\left(\tilde{\Sigma}^{-1} - \frac{1}{\iota'\tilde{\Sigma}^{-1}\iota}\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}\right)\right\|_F \leq \left\|\tilde{\Sigma}^{-1}\right\|_F + \frac{\iota'\tilde{\Sigma}^{-2}\iota}{\iota'\tilde{\Sigma}^{-1}\iota}. \tag{60}$$

The inequality holds after applying triangle equation and making use of the fact that

$$\left\|\tilde{\Sigma}^{-1} - \frac{1}{\iota'\tilde{\Sigma}^{-1}\iota}\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}\right\|_F = \left\|\tilde{\Sigma}^{-1} - \frac{1}{\iota'\tilde{\Sigma}^{-1}\iota}\alpha\alpha'\right\|_F, \tag{61}$$

where $\alpha := \tilde{\Sigma}^{-1}\iota$ and recognizing that for the rank 1 symmetric matrix $\|\alpha\alpha'\|_F = \|\alpha\|_2^2$. Now note that

$$\left\|\tilde{\Sigma}^{-1}\right\|_F = \left\|\left(I + \frac{\gamma}{\beta}\Sigma\right)^{-1}\right\|_F \tag{62}$$

$$= \sqrt{\sum_{i=1}^{N}\left(\frac{\beta}{\gamma}\right)^2\left(\frac{1}{\frac{\beta}{\gamma}+\lambda_i}\right)^2} \tag{63}$$

$$= \left(\frac{\beta}{\gamma}\right)\sqrt{\sum_{i=1}^{N}\left(\frac{1}{\frac{\beta}{\gamma}+\lambda_i}\right)^2}. \tag{64}$$

Also we can derive $\frac{1}{1+\frac{\gamma}{\beta}\lambda_1} < \frac{\iota\tilde{\Sigma}^{-2}\iota}{\iota'\tilde{\Sigma}^{-1}\iota} < \frac{1}{1+\frac{\gamma}{\beta}\lambda_N}$. The limits can be computed by decomposing $\frac{\gamma}{\beta}\Sigma$ into O orthogonal and Ξ diagonal. Then,

$$\iota'\tilde{\Sigma}^{-1}\iota = \iota'\left(I + \frac{\gamma}{\beta}\Sigma\right)^{-1}\iota = \kappa'\left(I+\Xi\right)^{-1}\kappa = \sum_{i=1}^{N}\frac{\kappa_i^2}{\frac{1}{1+\frac{\gamma}{\beta}\lambda_i}}, \tag{65}$$

where $\kappa = O\iota$. For the ratio $\frac{\iota\tilde{\Sigma}^{-2}\iota}{\iota'\tilde{\Sigma}^{-1}\iota}$ we get

$$\frac{\iota\tilde{\Sigma}^{-2}\iota}{\iota'\tilde{\Sigma}^{-1}\iota} = \frac{\sum_{i=1}^{N}\kappa_i^2/(\frac{1}{1+\frac{\gamma}{\beta}\lambda_i})^2}{\sum_{i=1}^{N}\kappa_i^2/(\frac{1}{1+\frac{\gamma}{\beta}\lambda_i})}. \tag{66}$$

The weights κ_i are constrained by $\|\kappa_i\|_2^2 = \|\iota\|^2 = N$, which results in the inequalities stated above. Putting everything together, we get

$$\left\|\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)\right\|_F \leq \left\|\tilde{\Sigma}^{-1}\right\|_F + \frac{\iota\tilde{\Sigma}^{-2}\iota}{\iota'\tilde{\Sigma}^{-1}\iota} \tag{67}$$

$$\leq \left(\frac{\beta}{\gamma}\right)\sqrt{\sum_{i=1}^{N}\left(\frac{1}{\frac{\beta}{\gamma}+\lambda_i}\right)^2} + \frac{1}{1+\frac{\gamma}{\beta}\lambda_N}. \tag{68}$$

The statement in the proposition holds, as for β reaching 0, the term $\frac{\beta}{\gamma}$ vanishes, $\sqrt{\sum_{i=1}^{N}\left(\frac{1}{\frac{\beta}{\gamma}+\lambda_i}\right)^2}$

converges towards a finite constant (as all eigenvalues of Σ are finite), and $\frac{1}{1+\frac{\gamma}{\beta}\lambda_N}$ also vanishes. Thus, $\exists\beta > 0$ such that for all $\tilde{\beta} \in [0, \beta) : \left\| \left(\frac{\tilde{\beta}}{\gamma} A(\Sigma^*) \right) \right\|_F < 1$. $\square$

Proof of Proposition 3. By exploiting that $\iota'\omega^*_{t+1} = 1$ and $g^{*\prime}\omega^*_{t+1} = ||\omega^*_{t+1} - \omega_{t^+}||_1$, $\Sigma_{\frac{\beta}{\gamma}}\omega^*_{t+1}$ can be written as

$$\Sigma_{\frac{\beta}{\gamma}}\omega^*_{t+1} = \Sigma\omega^*_{t+1} + \frac{\beta}{\gamma}g^* + \frac{\beta}{\gamma}||\omega^*_{t+1} - \omega_{t^+}||_1\iota \tag{69}$$

$$= \frac{\lambda}{\gamma}\iota + \frac{\beta}{\gamma}||\omega^*_{t+1} - \omega_{t^+}||_1\iota \tag{70}$$

$$= \frac{\beta}{\gamma}\left(\lambda + ||\omega^*_{t+1} - \omega_{t^+}||_1\right)\iota \tag{71}$$

Then, solving for ω^*_{t+1} yields

$$\omega^*_{t+1} \propto \Sigma_{\frac{\beta}{\gamma}}\iota. \tag{72}$$

Note that ω^*_{t+1} has the same direction as the solver of a 'classical' GMV problem $\omega^*_\beta := \Sigma^{-1}_{\frac{\beta}{\gamma}}\iota / \iota'\Sigma^{-1}_{\frac{\beta}{\gamma}}\iota$. The sum-up constraint completes the proof $\Rightarrow \omega^*_{t+1} = \omega_\beta$. $\square$

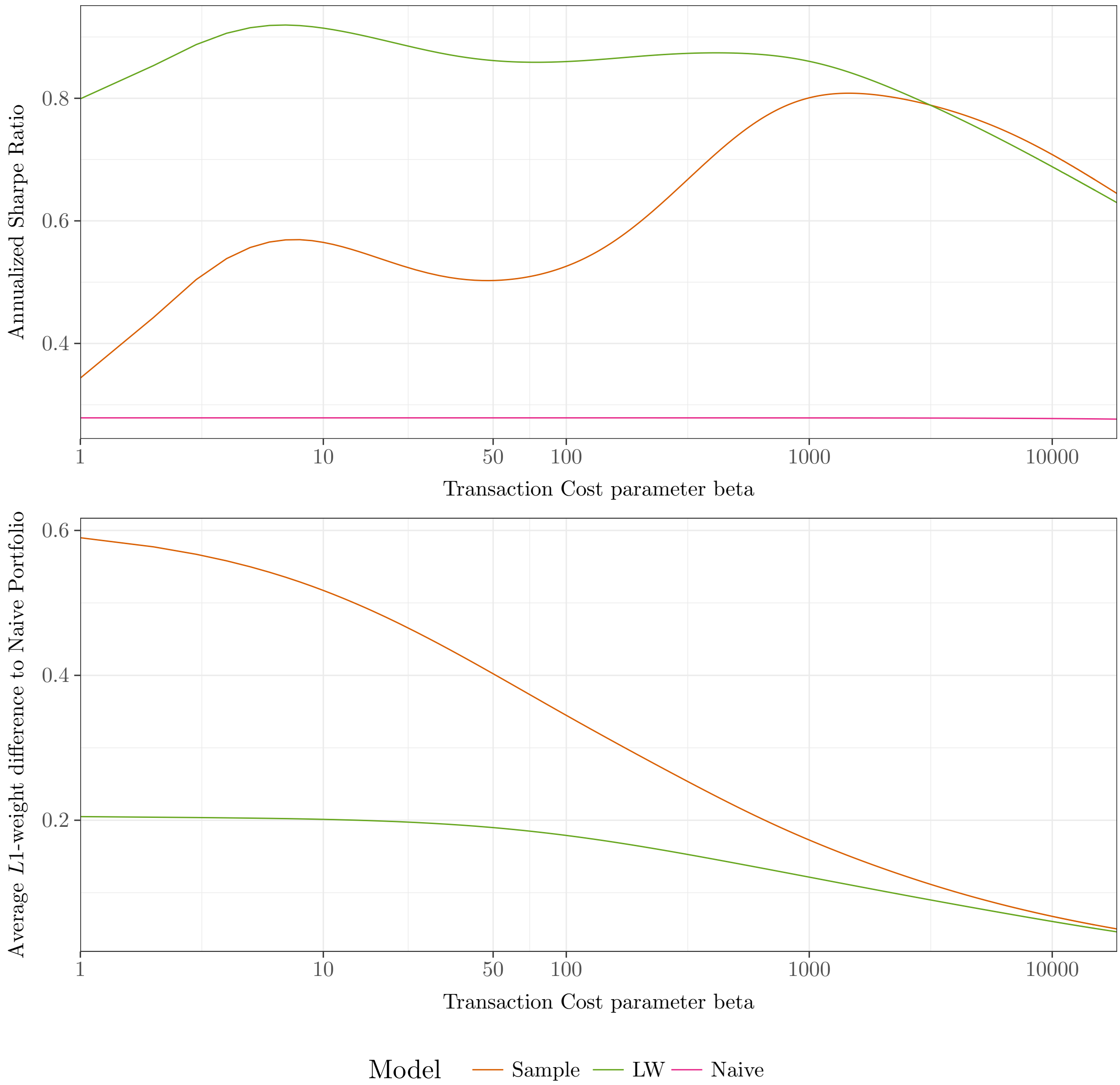


Figure 1: Annualized realized Sharpe ratio net of transaction costs for portfolios based on $N = 308$ assets with daily re-allocation for the period from June 2007 until March 2017 with quadratic transaction costs. The portfolio weights are computed as solutions to optimization problem (3) with varying choices of the transaction cost parameter β (in basis points). The variance-covariance matrix is estimated using the sample covariance matrix and a regularized version (Ledoit and Wolf, 2004) (both based on a rolling window of 500 days). The mean of the returns, μ_t, is set to 0. The second panel corresponds to the average L_1 difference of the estimated weights and the naive (equally weighted) portfolio.

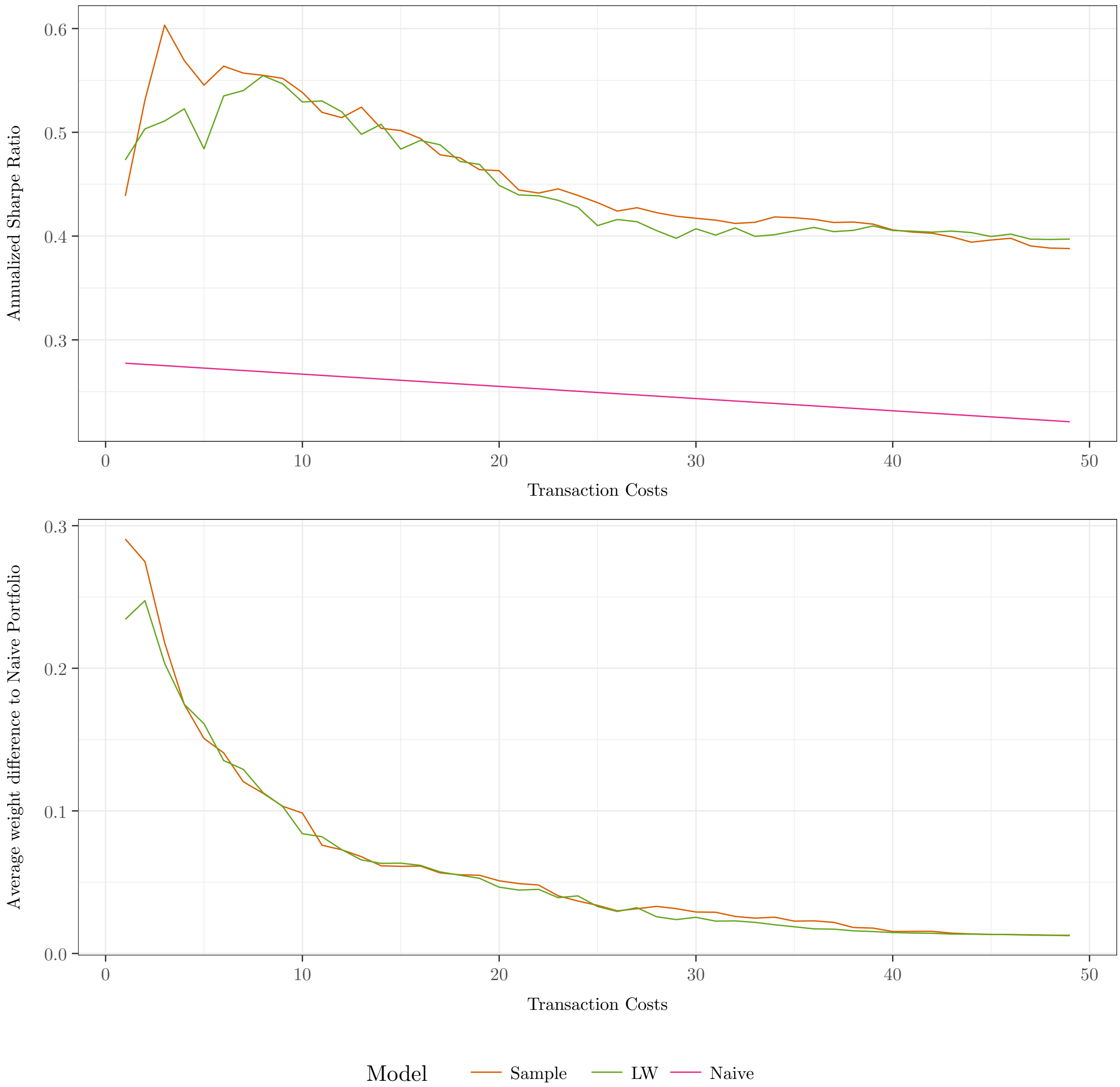
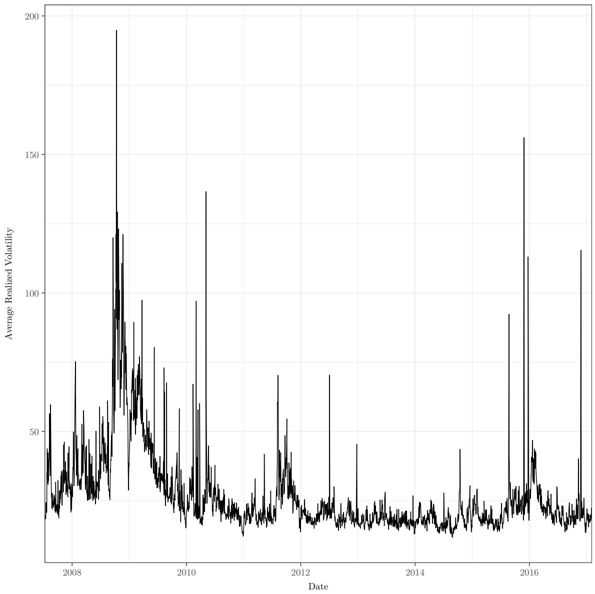
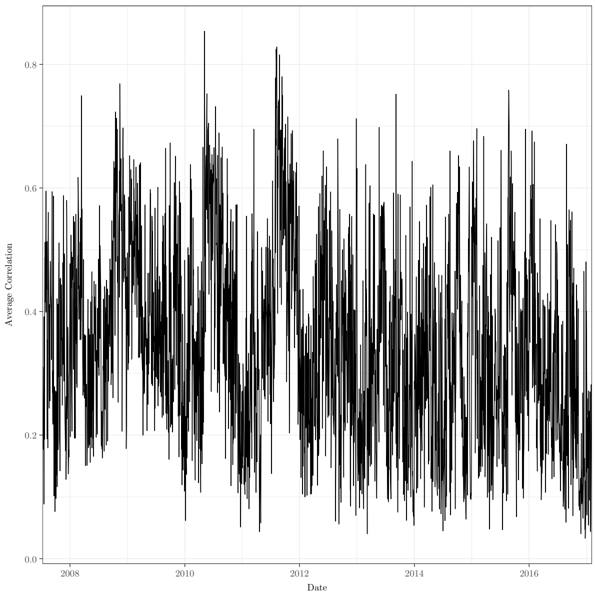


Figure 2: Annualized realized Sharpe ratio net of transaction costs for portfolios based on $N = 308$ assets with daily re-allocation for the period from June 2007 until March 2017 with L_1 transaction costs as in Equation (9). The portfolio weights are computed as solutions to optimization problem (10) with varying choices of the transaction cost parameter β (in basis points). The variance-covariance matrix is estimated using the sample covariance matrix and a regularized version (Ledoit and Wolf, 2004) (both based on a rolling window of 500 days). The mean of the returns, μ_t, is set to 0. The second panel corresponds to the average L_1 difference of the estimated weights and the naive (equally weighted) portfolio.



(a) Average Realized Volatilities



(b) Average Correlations

Figure 3: Cross-sectional averages of daily realized volatilities based on intra-daily data for $N = 308$ assets. Realized volatilities are computed using univariate realized kernels Barndorff-Nielsen et al. (2008). We report annualized volatilities in percentage points. The second panel denotes the mean of the off-diagonal elements of the estimated daily correlation based on blocked realized kernel matrices computed using 4 groups.

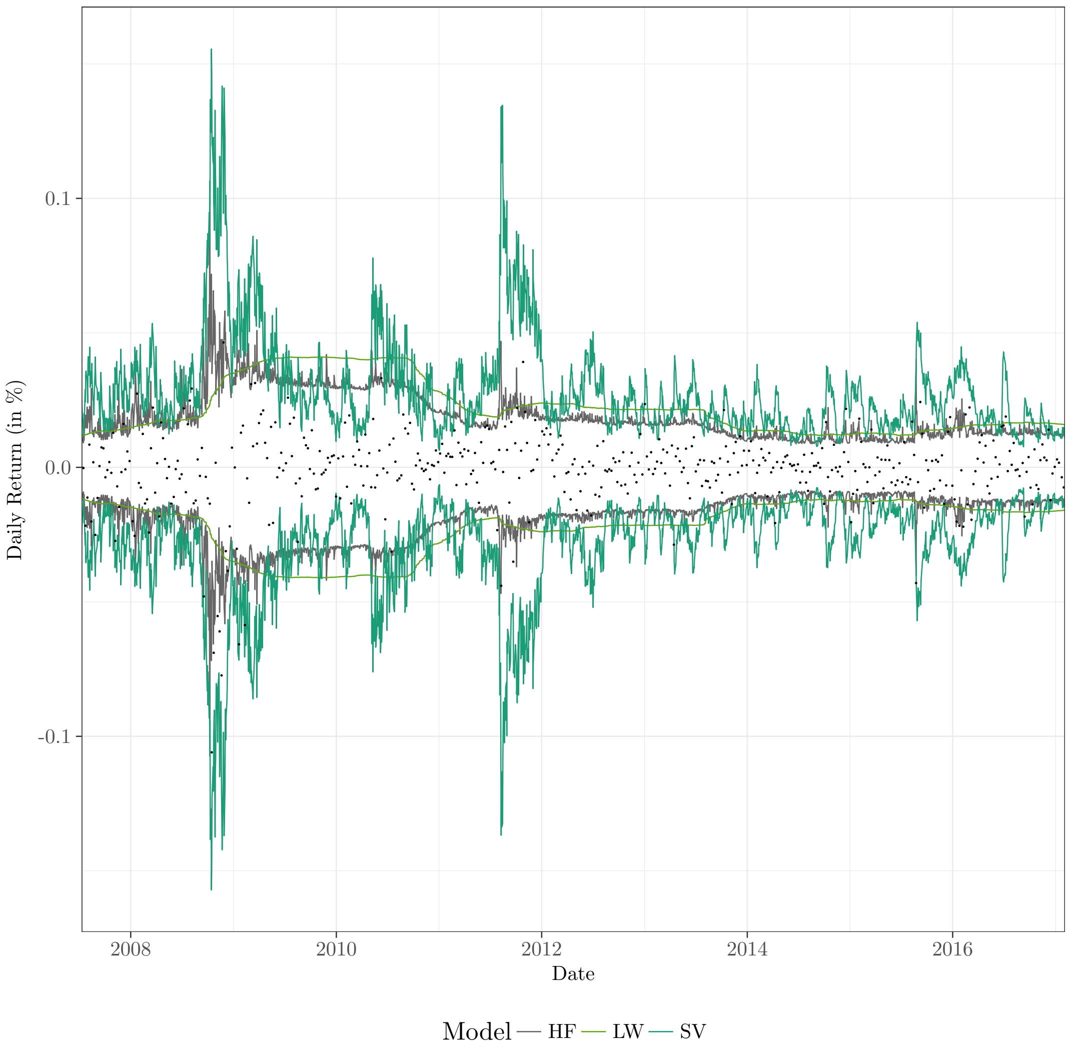


Figure 4: Predictive credible regions of the stochastic volatility factor model with $j = 3$ factors, the HF model with 4 liquidity groups and the Ledoit-Wolf (2004) shrinkage, computed as described in Table 1. For every model, a sample of the predicted portfolio return of the $1/N$ allocation is generated by first sampling from the predictive return distribution $r_{t+1}^{\mathcal{M}_k,(l)}$ and then evaluating the vector $\frac{1}{N}\iota' r_{t+1}^{\mathcal{M}_k,(l)}$. Lines indicate the 0.025 and 0.975 quantiles of the predictions. The dots indicate the observed (true) return of the naive portfolio at time $t + 1$.

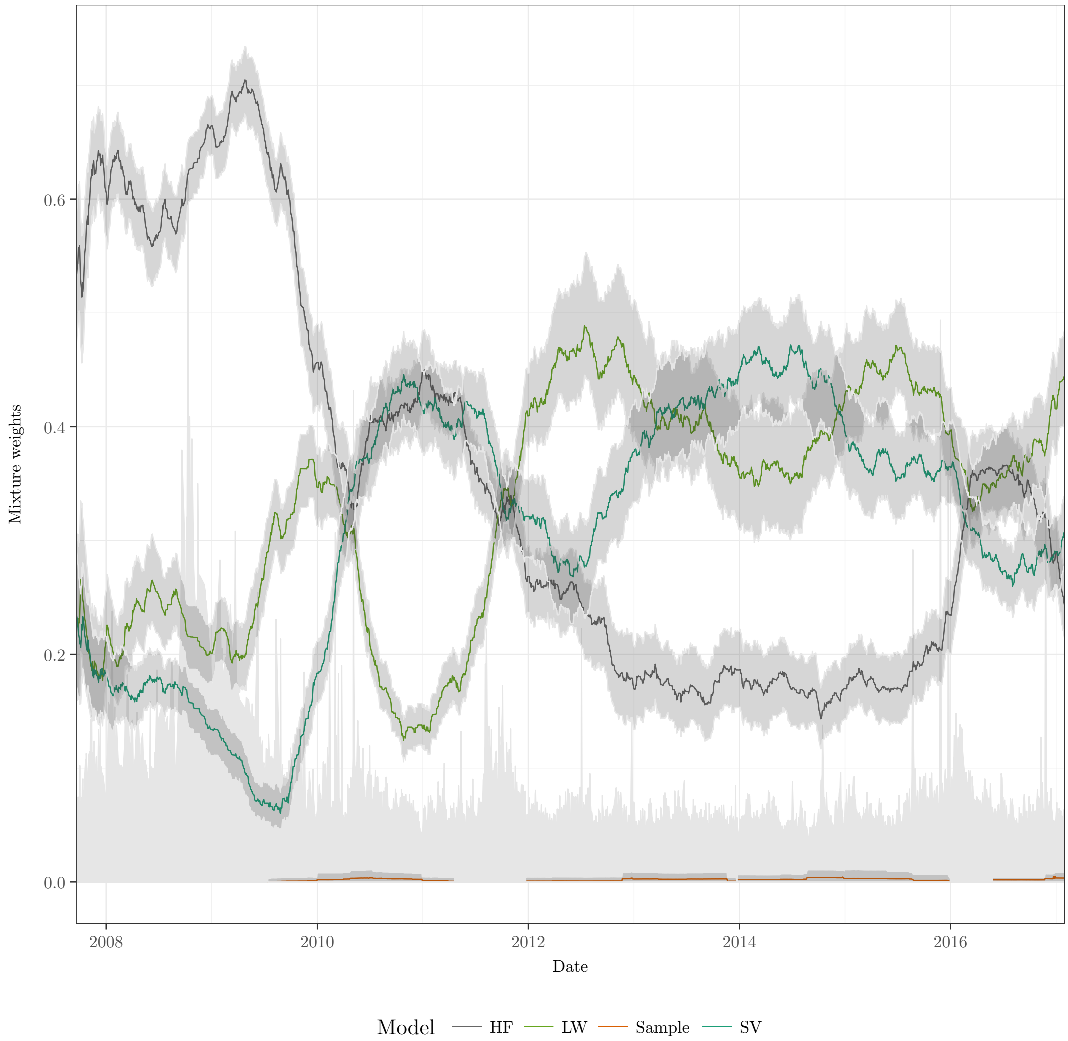


Figure 5: Combination weights computed based on (24). The weights are sequentially updated to maximize the log posterior predictive likelihood of past observed returns based on a rolling-window of 250 days. The lines correspond to the mean of the bootstrapped combination weights. The gray confidence bands denote the 2.5% and 97.5% quantiles of the bootstrapped combination weights. The shaded gray area in the background visualizes a scaled time series of the daily estimated volatility across all 308 markets and corresponds to the data used in Panel a) of Figure 3.

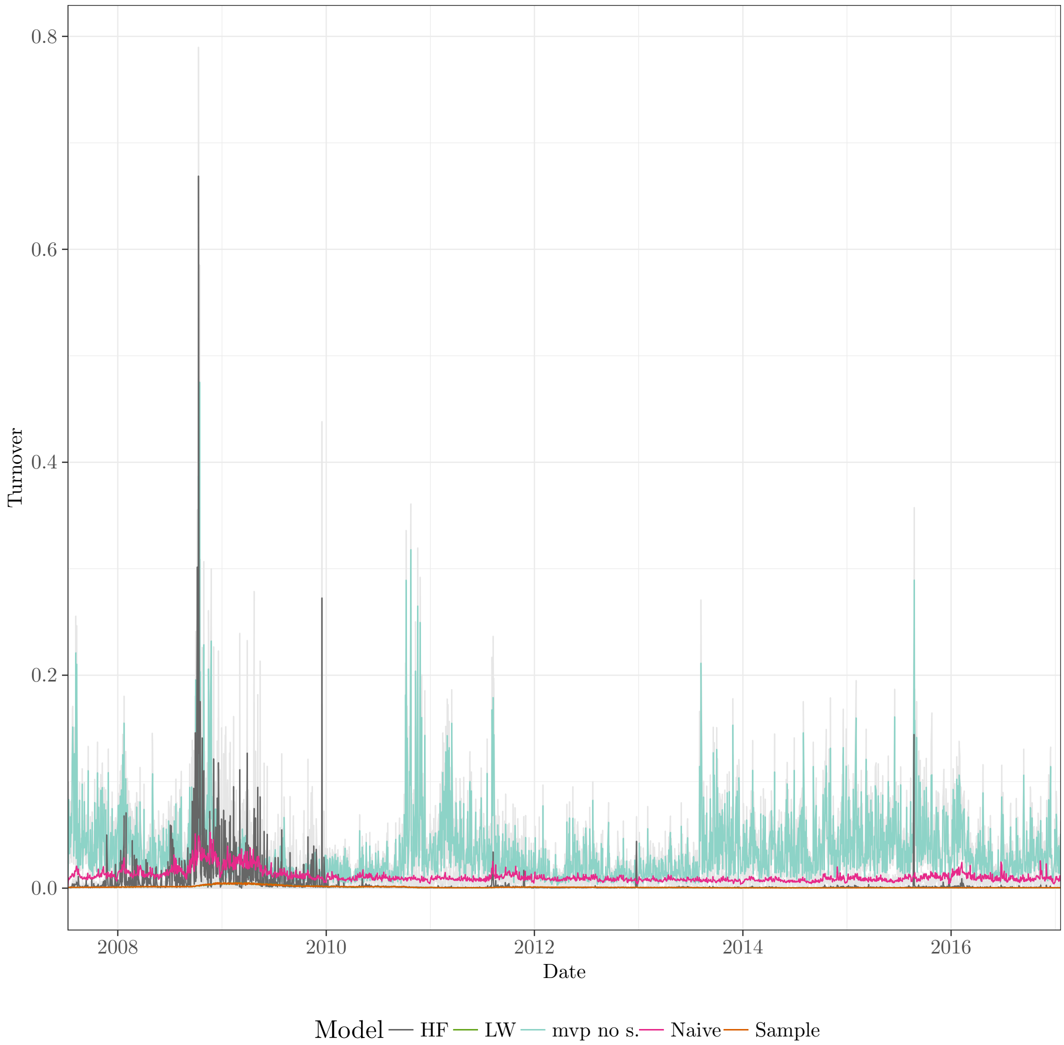


Figure 6: Time series of daily turnover implied by each strategy (TO^k), measured as the L_1 norm of daily re-balancing. The shaded areas correspond to bootstrapped confidence intervals.

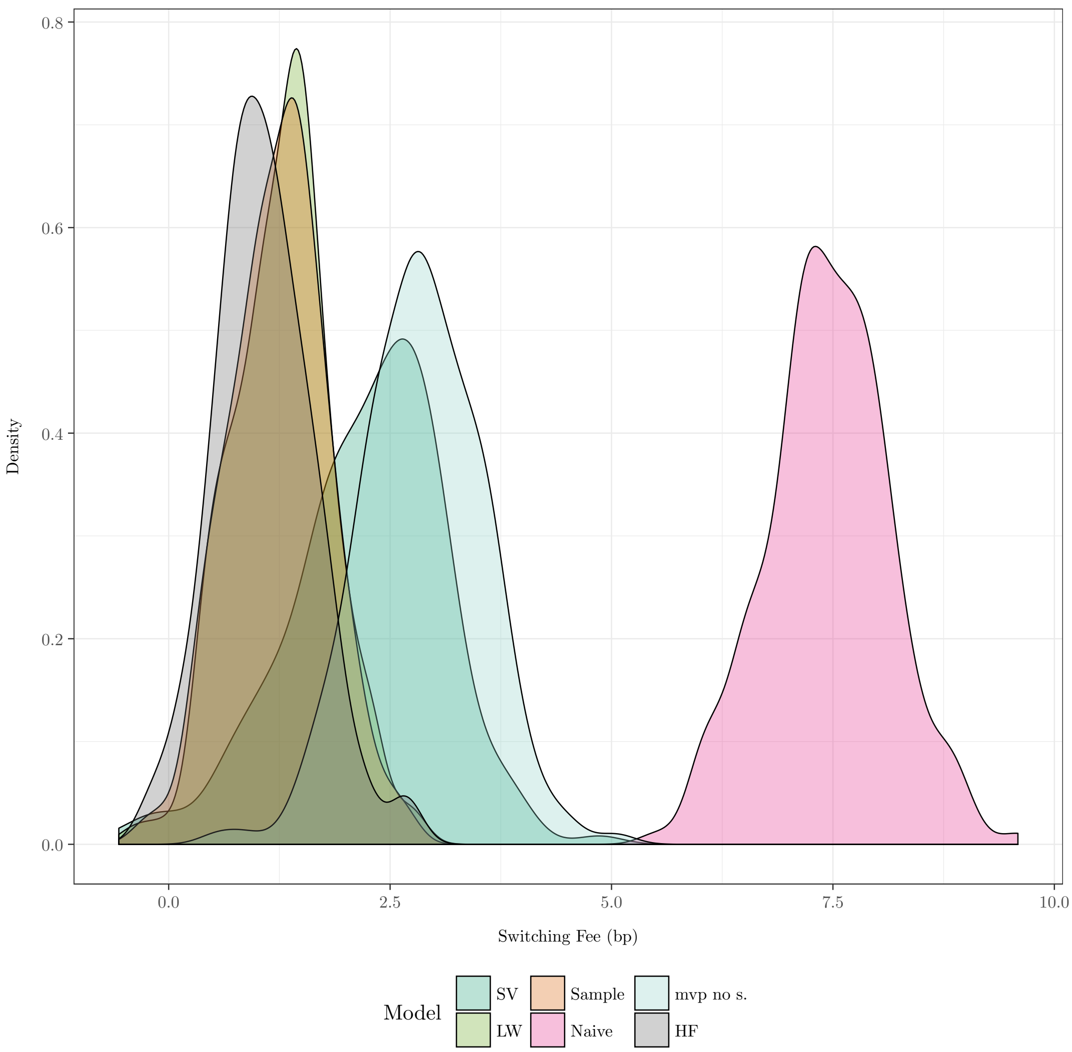


Figure 7: Bootstrapped distribution of annual switching fees in basis points for switching to the portfolio implied by model mixing. Computed for an investor with power utility and risk aversion parameter $\gamma = 4$ based on the 200 bootstrapped realized portfolio performances net of transaction costs.

CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

Recent Issues

All CFS Working Papers are available at *www.ifk-cfs.de*.

No.	Authors	Title
581	Martin D. Gould, Nikolaus Hautsch, Sam D. Howison, and Mason A. Porter	*Counterparty Credit Limits: An Effective Tool for Mitigating Counterparty Risk?*
580	Nikolaus Hautsch, Michael Noé, and S. Sarah Zhang	*The Ambivalent Role of High-Frequency Trading in Turbulent Market Periods*
579	Klaus Adam and Henning Weber	*Optimal Trend Inflation*
578	Günter Coenen, Michael Ehrmann, Gaetano Gaballo, Peter Hoffmann, Anton Nakov, Stefano Nardelli, Eric Persson, and Georg Strasser	*Communication of monetary policy in unconventional times*
577	Francis X. Diebold and Frank Schorfheide	*Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility*
576	Mert Demirer, Francis X. Diebold, Laura Liu, and Kamil Yilmaz	*Estimating global bank network connectedness*
575	Francis X. Diebold, Laura Liu, and Kamil Yilmaz	*Commodity Connectedness*
574	Annamaria Lusardi, Olivia S. Mitchell, and Noemi Oggero	*Debt and Financial Vulnerability on the Verge of Retirement*
573	Donald B. Keim and Olivia S. Mitchell	*Simplifying Choices in Defined Contribution Retirement Plan Design: A Case Study*