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

Nikolaus Hautsch   Stefan Voigt

Abstract

We theoretically and empirically study portfolio optimization under transaction costs and establish a link between turnover penalization and covariance shrinkage with the penalization governed by transaction costs. We show how the ex ante incorporation of transaction costs shifts optimal portfolios towards regularized versions of efficient allocations. The regulatory effect of transaction costs is studied in an econometric setting incorporating parameter uncertainty and optimally combining predictive distributions resulting from high-frequency and low-frequency data. In an extensive empirical study, we illustrate that turnover penalization is more effective than commonly employed shrinkage methods and is crucial in order to construct empirically well-performing portfolios.

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 rebalancing costly and thus sub-optimal.

Although there is a huge literature on the statistics of portfolio allocation, it 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, 2004, 2012) and Fan et al. (2008), 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, see, e.g., Barndorff-Nielsen and Shephard (2004). Another segment of the literature studies the effects of ignoring parameter uncertainty and model uncertainty arising from changing market regimes and structural breaks.[1] Further literature is devoted to the role of transaction costs in portfolio allocation strategies. In the presence of transaction costs, the benefits of reallocating wealth may be smaller than the costs associated with turnover. This aspect has been investigated theoretically, among others, for one risky asset by Magill and Constantinides (1976) and Davis and Norman (1990). Subsequent extensions to the case with multiple assets have been proposed by Taksar et al. (1988), Akian et al. (1996), Leland (1999) and Balduzzi and Lynch (2000). More recent papers on empirical approaches which explicitly account for transaction costs include Liu (2004), Lynch and Tan (2010), Gârleanu and Pedersen (2013) and DeMiguel et al. (2014, 2015).

Our paper connects the work on shrinkage estimation and transaction costs in two ways: First, we show a close connection between covariance regularization and the effects of transaction costs in the optimization procedure.[2] Second, we empirically document these effects in a large-scale study mimicking portfolio optimization under preferably realistic conditions based on a large panel of assets through more than 10 years.

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.[3] In financial practice, however, the costs of portfolio rebalancing are taken into account *ex ante* and thus are part of the optimization problem. Thus, our objective is to understand the

---

1. The effect of ignoring estimation uncertainty is considered, among others, by Brown (1976), Jobson and Korkie (1980), Jorion (1986) and Chopra and Ziemba (1993). Model uncertainty is investigated, for instance, by Wang (2005), Garlappi et al. (2007) and Pflug et al. (2012).

2. The work of DeMiguel and Olivares-Nadal (2018) is close in spirit to our approach. Though both papers derive optimal portfolios after imposing a L_p penalty on rebalancing, the implications, however, are different. While we focus on the regularization effect, DeMiguel and Olivares-Nadal (2018) point out the close relationship to a robust Bayesian decision problem where the investor imposes priors on her optimal portfolio.

3. See, e.g., DeMiguel et al. (2009) or Hautsch et al. (2015).

effect of turnover penalization on the ultimate object of interest of the investor – the optimal portfolio allocation. This focus is notably different from the aim of providing sensible estimates of asset return covariances (via regularization methods), which are then plugged into a portfolio problem. Instead, we show how the presence of transaction costs changes the optimal portfolio and provide an alternative interpretation in terms of parameter shrinkage. In particular, we illustrate that *quadratic* transaction costs can be interpreted as shrinkage of the variance-covariance matrix towards a diagonal matrix and a shift of the mean that is proportional to transaction costs and current holdings. Transaction costs *proportional* to the amount of rebalancing imply a regularization of the covariance matrix, acting similarly as the least absolute shrinkage and selection operator (Lasso) by Tibshirani (1996) in a regression problem and imply to put more weight on a buy-and-hold strategy. The regulatory effect of transaction costs results in better conditioned covariance estimates and significantly reduces the amount (and frequency) of rebalancing. These mechanisms imply strong improvements of portfolio allocations in terms of expected utility and Sharpe ratios compared to the case where transaction costs are neglected.

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 rebalancing. 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 and Amisano (2011). This makes the setup sufficiently flexible to utilize a long sample covering high-volatility and low-volatility periods subject to obvious structural breaks. As a by-product we provide insights into the time-varying nature of the predictive ability of individual models under transaction costs and to which extent suitable forecast combinations may result into better portfolio allocations.

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 to integrate out parameter uncertainty and to construct posterior predictive asset return distributions based on time-varying mixtures making use of Bayesian computation techniques. 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-inverse Wishart mixture close to 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).[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 (2003, 2004).

To our best knowledge, this paper provides the first study evaluating the predictive power of high-frequency and low-frequency models in a large-scale portfolio framework under such generality, while utilizing data of 2,409 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.

We can summarize the following findings: First, none of the underlying 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 rebalancing. This result changes drastically when transaction costs are considered in the optimization *ex ante*. Second, when incorporating transaction costs into the optimization problem, performance differences between competing predictive models for the return distribution become smaller. It is shown that none of the underlying 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. 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. Fourth, naive strategies, variants of minimum variance allocations and further competing 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. All proofs, more detailed information on the implementation of the estimation procedures described in the

---

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.

paper as well as additional results are provided in an accompanying Online Appendix.

2 The Role of Transaction Costs

2.1 Decision Framework

We consider an investor equipped with a utility function $U_\gamma(r)$ depending on returns r and risk aversion parameter γ. 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 $r_{t,i} := \frac{p_{t,i} - p_{t-1,i}}{p_{t-1,i}}$ are the simple (net) 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 rebalancing. 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 vector $\Delta_t := \omega_{t+1} - \omega_{t+}$ between the new allocation ω_{t+1} and the allocation right before readjustment, $\omega_{t+} := \frac{\omega_t \circ (1+r_t)}{1+\omega_t' r_t}$, where the operator $\circ$ denotes element-wise multiplication. The vector ω_{t+} builds on the 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 her current beliefs on the distribution of returns $p_t(r_{t+1}|\mathcal{D}) := p(r_{t+1}|R_t, \mathcal{H}_t, \mathcal{D})$ of the next period and the current portfolio weights ω_{t+}:

solve numerically (approx.)

$$\omega_{t+1}^* := \underset{\omega \in \mathbb{R}^N, \, \iota'\omega=1}{\arg\max} \; E\left(U_\gamma\left(\omega'(1+r_{t+1}) - \nu_t(\omega)\right)|R_t, \mathcal{H}_t, \mathcal{D}\right)$$

$$= \underset{\omega \in \mathbb{R}^N, \, \iota'\omega=1}{\arg\max} \int_{\mathbb{R}^N} U_\gamma\left(\omega'(1+r_{t+1}) - \nu_t(\omega)\right) p_t(r_{t+1}|\mathcal{D}) \mathrm{d}r_{t+1} \qquad \text{(EU)}$$

where ι is a vector of ones. Note that optimization problem (EU) reflects the problem of an investor who constantly monitors her portfolio and exploits all available information, but rebalances only if the costs implied by deviations from the path of optimal allocations exceed the costs of rebalancing. This form of myopic portfolio optimization ensures optimality (after transaction costs) of allocations at each point in time.[5] Accordingly, the optimal wealth allocation ω_{t+1}^* from representation (EU) is

---

5. The investment decision is myopic in the sense that it does not take into account the opportunity to rebalance the portfolio next day. Thus, hedging demand for assets is not adjusted for. Although attempts exist to investigate

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 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

$$\nu_{L_2}(\omega_{t+1}, \omega_{t^+}, \beta) := \frac{\beta}{2}(\omega_{t+1} - \omega_{t^+})'(\omega_{t+1} - \omega_{t^+}) \tag{2}$$

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

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

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 (ii) the mean is shifted by $\beta\omega_{t^+}$.

The shift from μ to $\mu^* = \mu + \beta\omega_{t^+}$ can be alternatively interpreted by exploiting $\omega'\iota = 1$ and reformulating the problem as

$$\omega_{t+1}^* = \underset{\omega \in \mathbb{R}^N, \iota'\omega=1}{\arg\min} \; \frac{\gamma}{2}\omega'\Sigma^*\omega - \omega'\left(\mu + \beta\left(\omega_{t^+} - \frac{1}{N}\iota\right)\right). \tag{6}$$

---

the effect of (long-term) strategic decisions of the investor, numerical solutions are infeasible in our high-dimensional asset space (see, e.g. Campbell et al. (2003)). Further, under certain restrictions regarding the utility function of the investor, the myopic portfolio choice indeed represents the optimal solution to the (dynamic) control problem.

The shift of the mean vector is therefore 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.

Hence, if β increases, Σ^* is shifted towards a diagonal matrix representing the case of uncorrelated assets. The regularization effect of Σ^* exhibits some similarity to the implications of the shrinkage approach proposed by Ledoit and Wolf (2003, 2004). While the latter propose to shrink the eigenvalues of the covariance matrix Σ, (4) implies shifting all eigenvalues by a constant $\frac{\beta}{\gamma}$ which has a stabilizing effect on the conditioning of Σ^*. As shown in Section A.1 of the Online Appendix, a direct link to (linear) shrinkage can be constructed by imposing asset-specific transaction costs of the form $v(\omega_{t+1}, \omega_{t+}, B) := (\omega_{t+1} - \omega_{t+})' B (\omega_{t+1} - \omega_{t+})$ for a positive definite matrix B. Then, the linear shrinkage estimator of Ledoit and Wolf (2003, 2004) originate as soon as transaction costs are specifically related to the eigenvalues of Σ.

Note, however, that in our context, transaction costs β are treated as an exogenous parameter, which is determined by the institutional environment. Thus it is not a 'free' parameter and is independent of Σ. This makes the transaction-cost-implied regularization different to the case of linear shrinkage and creates a seemingly counterintuitive result (at first sight): Due to the fact that transaction costs β are independent of Σ, periods of high volatility (i.e., an increase of Σ) imply a relatively lower weight on the identity matrix I and thus less penalization compared to the case of a period where Σ is scaled downwards. This is opposite to the effect which is expected for shrinkage, where higher volatility implies higher parameter uncertainty and thus stronger penalization. As shown by the lemma below, however, the impression of counterintuition disappears if one turns attention not to the covariance matrix *per se*, but to the objective of portfolio optimization.

Lemma 1. *Assume a regime of high volatility with $\Sigma^h = (1 + h)\Sigma$, where $h > 0$ and Σ is the asset return covariance matrix during calm periods. Assume $\mu = 0$. Then, the optimal weight ω_{t+1}^* is equivalent to the optimal portfolio based on Σ and (reduced) transaction costs $\frac{\beta}{1+h} < \beta$.*

Proof. See Section A.2 in the Online Appendix. ☐

Lemma 1 implies that during periods with high volatility (and covariances) the optimal portfolio decision is triggered *less* by rebalancing costs, but *more* by the need to reduce portfolio volatility. The consequence is a shift towards the (global) minimum-variance portfolio with weights $\omega \propto \left(\frac{\beta}{(1+h)\gamma} I + \Sigma \right)^{-1} \iota$. Alternatively, one can associate this scenario with a situation of higher risk aversion and thus a stronger need for optimal risk allocation.

The view of treating β as an exogenous transaction cost parameter can be contrasted to recent approaches linking transaction costs to volatility. In fact, if transaction costs are assumed to be *proportional* to volatility, $v(\omega_{t+1}, \omega_{t+}, \beta\Sigma) = \frac{\beta}{2} \Delta_t' \Sigma \Delta_t$, as, e.g., suggested by Gârleanu and Pedersen

(2013), we obtain $\Sigma^* = \left(1 + \frac{\beta}{\gamma}\right)\Sigma$. As shown in Lemma 2 in Section A.2 in the Online Appendix, in this case, the optimal weights take the form $\omega_{t+1}^{\beta\Delta} = \omega_{t+1}^{\gamma+\beta} + \frac{\beta}{\beta+\gamma}(\omega_{t+} - w^{\text{mvp}})$, where $\omega_{t+1}^{\gamma+\beta}$ is the efficient portfolio *without* transaction costs, parameters Σ and μ and risk aversion parameter $\gamma+\beta$, and $\omega^{\text{mvp}} := \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota$ corresponds to the minimum variance allocation. Hence, the optimal weights are a linear combination of $\omega_{t+1}^{\gamma+\beta}$ with weight 1 and of $(\omega_{t+} - w^{\text{mvp}})$ with weight $\frac{\beta}{\beta+\gamma}$. Changes in Σ thus only affect $\omega_{t+1}^{\gamma+\beta}$ and w^{mvp} but *not* the weight $\frac{\beta}{\beta+\gamma}$, which might be interpreted as a 'shrinkage intensity' (in a wider sense). Hence, the regulatory effect of *endogenous* transaction costs can be ambiguous and very much depends on the exact specification of the link between β and Σ. In contrast, *exogenous* transaction costs imply a distinct regulatory effect, which is shown to be empirically successful as documented in Section 5.

Focusing on $v_{L_2}(\omega_{t+1}, \omega_{t+}, \beta)$ allows to us to investigate the relative importance of the current holdings on the optimal financial decision. Proposition 1 shows the effect of rising transaction costs on optimal rebalancing.

Proposition 1.

$$\lim_{\beta\to\infty} \omega_{t+1}^* = \left(I - \frac{1}{N}\iota\iota'\right)\omega_{t+} + \frac{1}{N}\iota = \omega_{t+}. \tag{7}$$

Proof. See Section A.2 in the Online Appendix. □

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,

$$\omega(\mu,\Sigma) := \frac{1}{\gamma}\left(\underbrace{\Sigma^{-1} - \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota'\Sigma^{-1}}_{:=A(\Sigma)}\right)\mu + \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota. \tag{8}$$

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

$$\omega_T = \sum_{i=0}^{T-1}\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^i \omega(\mu,\Sigma^*) + \left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^T \omega_0. \tag{9}$$

Therefore, 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. Proposition 2 states that for $\beta \to \infty$, the long-run optimal portfolio ω_T is driven only by the initial allocation ω_0.

Proposition 2. $\lim_{\beta\to\infty} \omega_T = \omega_0$.

7

Proof. See Section A.2 in the Online Appendix. □

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 3. $\exists \beta^* > 0 \; \forall \beta \in [0, \beta^*) : \left\| \left(\frac{\beta}{\gamma} A(\Sigma^*) \right) \right\|_F < 1$, *where* $\| \cdot \|_F$ *denotes the Frobenius norm* $\|A\|_F := \sqrt{\sum\limits_{i=1}^{N} \sum\limits_{i=1}^{N} A_{i,j}^2}$ *for an* $N \times N$ *matrix* A.

Proof. See Section A.2 in the Online Appendix. □

Using Proposition 3 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 Proposition 4 we show that in the long run, we obtain convergence towards the efficient portfolio:

Proposition 4. *For* $T \to \infty$ *and* $\beta < \beta^*$ *the series* ω_T *converges to a unique fix-point given by*

$$\omega_\infty = \left(I - \frac{\beta}{\gamma} A(\Sigma^*) \right)^{-1} \omega(\mu, \Sigma^*) = \omega(\mu, \Sigma). \tag{10}$$

Proof. See Section A.2 in the Online Appendix. □

Note, that the location of the initial portfolio ω_0 itself does not play a role on the upper threshold β^* ensuring 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

Though attractive from an analytical perspective, transaction costs of quadratic form may represent an unrealistic proxy of costs associated with trading in real financial markets, see, among others, Tóth et al. (2011) and Engle et al. (2012). Instead, in the literature, there is widespread use of transaction cost measures proportional to the sum of absolute rebalancing (L_1-norm of rebalancing), which impose a stronger penalization on turnover and are more realistic.[6] Transaction costs proportional to the L_1-norm of rebalancing $\Delta_t = \omega_{t+1} - \omega_{t^+}$ yield the form

$$\nu_{L_1}(\omega_{t+1}, \omega_{t^+}, \beta) := \beta \|\omega_{t+1} - \omega_{t^+}\|_1 = \beta \sum_{i=1}^{N} |\omega_{t+1,i} - \omega_{t^+,i}|, \tag{11}$$

with cost parameter $\beta > 0$. Although the effect of proportional transaction costs on the optimal portfolio cannot be derived in closed-form comparable to the quadratic (L_2) case discussed above,

---

6. Literature typically employs a penalization terms of 50 bp, see, for instance, DeMiguel et al. (2009) and DeMiguel and Olivares-Nadal (2018).

the impact of turnover penalization can be still interpreted as a form of regularization. If we assume for simplicity of illustration $\mu = 0$, the optimization problem (1) corresponds to

$$\omega_{t+1}^* := \underset{\omega \in \mathbb{R}^N, \ \iota'\omega=1}{\arg\min} \frac{\gamma}{2}\omega'\Sigma\omega + \beta \left\| \omega_{t+1} - \omega_{t+} \right\|_1 \tag{12}$$

$$= \underset{\Delta_t \in \mathbb{R}^N, \ \iota'\Delta_t=0}{\arg\min} \gamma\Delta_t'\Sigma\omega_{t+} + \frac{\gamma}{2}\Delta_t'\Sigma\Delta_t + \beta \left\| \Delta_t \right\|_1. \tag{13}$$

The first-order conditions of the constrained optimization are

$$\gamma\Sigma \underbrace{(\Delta_t + \omega_{t+})}_{\omega_{t+1}^*} + \beta\tilde{g} - \lambda\iota = 0, \tag{14}$$

$$\Delta_t'\iota = 0, \tag{15}$$

where $\tilde{g}$ is the vector of sub-derivatives of $\left\| \omega_{t+1} - \omega_{t+} \right\|_1$, i.e., $\tilde{g} := \partial \left\| \omega_{t+1} - \omega_{t+} \right\|_1 / \partial\omega_{t+1}$, 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

$$\omega_{t+1}^* = \left(1 + \frac{\beta}{\gamma}\iota'\Sigma^{-1}\tilde{g} \right)\omega^{\text{mvp}} - \frac{\beta}{\gamma}\Sigma^{-1}\tilde{g}, \tag{16}$$

where $\omega^{\text{mvp}} := \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota$ corresponds to the weights of the GMV portfolio. Proposition 5 shows that this optimization problem can be formulated as a standard (regularized) minimum variance problem.

Proposition 5. *Portfolio optimization problem (12) is equivalent to the minimum variance problem with*

$$\omega_{t+1}^* = \underset{\omega \in \mathbb{R}^N, \ \iota'\omega=1}{\arg\min} \omega'\Sigma_{\frac{\beta}{\gamma}}\omega, \tag{17}$$

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

Proof. See Section A.2 in the Online Appendix. $\square$

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 rebalanced in the same direction. The result shows some similarities to Fan et al. (2012), who illustrate that the risk minimization problem with constrained weights

$$\omega_{t+1}^\vartheta = \underset{\omega \in \mathbb{R}^N, \ \iota'\omega=1, \ ||\omega||_1 \leq \vartheta}{\arg\min} \omega'\Sigma\omega \tag{18}$$

can be interpreted as the minimum variance problem

$$\omega_{t+1}^{\vartheta} = \underset{\omega \in \mathbb{R}^N, \, \iota'\omega=1}{\arg\min} \ \omega'\Sigma^{\vartheta}\omega, \tag{19}$$

with $\Sigma^{\vartheta} := \Sigma + \lambda\,(g\iota' + \iota g')$, 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 (18). Note, however, that in our case, 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 4 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 from implementing 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 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.[7] The unknown covariance matrix Σ_t is estimated in two ways: We compute the sample covariance estimator S_t and the shrinkage estimator $\hat{\Sigma}^{t,\text{Shrink}}$ by Ledoit and Wolf (2003, 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, portfolio weights are rebalanced 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 $r_t^{\beta,\text{nTC}} := r_t^{\beta} - \beta \left\| \omega_{t+1}^{\beta} - \omega_{t+}^{\beta} \right\|_2$.

Figure 1 displays annualized Sharpe ratios, computed as the ratio of the annualized sample mean and standard deviation of $r_t^{\beta,\text{nTC}}$, 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. As a benchmark, we provide the corresponding statistics for the naive portfolio without any rebalancing (buy and hold) and a naive version which reallocates weights back to $\frac{1}{N}\iota$ on a daily basis. We report values of β ranging from 1bp to $1,000,000$bp. To provide some guidance for the interpretation of these values, note that, for instance, for the naive allocation with daily readjustment, the adjusted

---

7. A description of the dataset and the underlying estimators is given in more detail in Section 5.1.

returns after transaction costs are

$$r_{t+1}^{\text{Naive, nTC}} := r_{t+1}^{\text{Naive}} - \beta \left\| \omega_{t+1} - \omega_{t^+} \right\|_2 = \bar{r}_{t+1} - \beta \frac{1}{N} \frac{1}{(1 + \bar{r}_{t+1})^2} \bar{\sigma}_{t+1}^2, \tag{20}$$

where $\bar{r}_t$ is the day t average simple return across all N assets and $\bar{\sigma}_t^2 = 1/N \sum_{i=1}^{N} (r_{t,i} - \bar{r}_t)^2$. Using sample estimates based on our data, penalizing turnover such that average daily returns of the naive portfolio are reduced by one basis point (which we consider very strict for the naive allocation), β should be at least in the magnitude of $3,000$ bp.[8] These considerations show that scenarios with $\beta < 100$ can be associated with rather small transaction costs.

Nonetheless, we observe that the naive portfolio is clearly outperformed for a wide range of values of β. 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 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 each of the 308 eigenvalues 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 covariance matrix, which adjusts ex-ante for transaction costs of $\beta = 1bp$, leads to a net-of-transaction-costs Sharpe ratio of 0.51, whereas neglecting transaction costs in the optimization yields a Sharpe ratio of only 0.21. 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 10bp 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.

For higher values of β we observe a similar pattern: Here, a decline in rebalancing 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 performance of the buy and hold strategy (with initial naive allocation $1/N$). This is reflected by the second panel of Figure 1, showing the average L_1 distance to the buy and hold portfolio weights.

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 penaliza-

---

8. In our sample ($N = 308$), the average daily return $\bar{r}$ is roughly $3bp$. The average variance $\bar{\sigma}^2$ is of the same magnitude.

tion 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. The third panel of Figure 1 illustrates the average turnover (in percent). Clearly, with increasing β, trading activity declines as shifting wealth becomes less desirable. Hence, we observe that differences in the out-of-sample performance between the two approaches decline if the turnover regularization becomes stronger.

Our motivation for reporting Sharpe ratios in Figure 1 originates from the fact that the Sharpe ratio is a very common metric for portfolio evaluation in practice. Note, however, that the maximization of the Sharpe ratio is *not* guaranteed by our underlying objective function (EU) that maximizes expected utility. Recall that in the latter framework, for a given β, allocation ω_{t+1}^* is optimal. If, however, one leaves this ground and aims at a different objective, such as, e.g., the maximization of the Sharpe ratio, one may be tempted to view β as a free parameter which needs to be chosen such that it maximizes the out-of-sample performance in a scenario where turnover with transaction costs $v_t\left(\omega_{t+1}, \omega_{t+}, \tilde{\beta}\right)$ is penalized *ex post*. Hence, in such a setting it might be useful to distinguish between an (endogenously chosen) *ex-ante* parameter β which differs from the (exogenously) given transaction cost parameter $\tilde{\beta}$ used in the *ex-post* evaluation. In Section A.3 of the Online Appendix we provide some evidence for this effect: Instead of computing the weights using the theoretically optimal penalty factor β, we compute Sharpe ratios for weights based on *sub-optimal ex-ante* $\tilde{\beta}$s. The results in Table A.1 show that small values for an *ex-ante* β are sufficient to generate high Sharpe ratios, even if the *ex-post* $\tilde{\beta}$ chosen for the evaluation is much higher. Furthermore, the table suggests that choosing $\beta = \tilde{\beta}$ does not necessarily provide the strongest empirical performance: In our setup, $\beta = 25bp$ leads to the highest Sharpe ratios even if $\tilde{\beta}$ is much higher.

Repeating the analysis based on proportional (L_1) transaction costs delivers qualitatively similar results. For low values of β, the Sharpe ratio increases in β as the effects of covariance regularization and a reduction of turnover overcompensate the effect of declining portfolio returns (after transaction costs). Overall, the effects, however, are less pronounced than in the case of quadratic transaction costs.[9]

3 Basic Econometric Setup

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

---

9. Note that L_1 transaction costs imply a regularization which acts similarly to a Lasso penalization. Such a penalization implies a strong dependence of the portfolio weights on the previous day's allocation. This affects our evaluation as the paths of the portfolio weights may differ substantially over time if the cost parameter β is slightly changed. A visualization of the performance for the L_1 case similarly to Figure 1 is provided in Section A.4 of the Online Appendix (Figure A.1).

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 $\mathcal{L}(r_t|\Theta, \mathcal{H}_t, \mathcal{M}_k)$ 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, see, e.g. Kan and Zhou (2007). Consequently, the resulting portfolio weights would be sub-optimal.

To accommodate parameter uncertainty and to impose 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 $\mathcal{L}(r_t|\Theta, \mathcal{H}_t, \mathcal{M}_k)$ and choosing a prior distribution $\pi(\Theta)$, the posterior distribution

$$\pi(\Theta|R_t, \mathcal{H}_t, \mathcal{M}_k) \propto \mathcal{L}(R_t|\Theta, \mathcal{H}_t, \mathcal{M}_k)\pi(\Theta) \tag{21}$$

reflects beliefs about the distribution of the parameters after observing the set of available information, $(R_t, \mathcal{H}_t)$. The (posterior) predictive distribution of the returns is then given by

$$r_{\mathcal{M}_k,t+1} \sim p(r_{t+1}|R_t, \mathcal{H}_t, \mathcal{M}_k) := \int \mathcal{L}(r_{t+1}|\Theta, \mathcal{H}_t, \mathcal{M}_k)\pi(\Theta|R_t, \mathcal{H}_t, \mathcal{M}_k)d\Theta. \tag{22}$$

If the precision of the parameters estimates is low, the posterior distribution $\pi(\Theta|R_t, \mathcal{H}_t, \mathcal{M}_k)$ yields a predictive return distribution with more mass in the tails than focusing only on $\mathcal{L}r_{t+1}|\hat{\Theta}, \mathcal{H}_t, \mathcal{M}_k$.

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.[10] Stacking the predictive distributions yields

$$r_{\mathcal{D},t+1}^{\text{vec}} := \text{vec}\left(\{r_{\mathcal{M}_1,t+1}, \ldots, r_{\mathcal{M}_K,t+1}\}\right) \in \mathbb{R}^{NK \times 1}. \tag{23}$$

The joint predictive distribution $p_t(r_{t+1}|\mathcal{D})$ is computed conditionally 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

---

10. Model combination in the context of return predictions has a long tradition in econometrics, starting from Bates and Granger (1969). See also Stock and Watson (1999, 2002); Weigend and Shi (2000); Bao et al. (2007) and Hall and Mitchell (2007). In finance, Avramov (2003) and Uppal and Wang (2003), among others, apply model combinations and investigate the effect of model uncertainty on financial decisions.

values and add up to one,

$$c_t \in \Delta_{[0,1]^K} := \left\{ c \in \mathbb{R}^K : c_i \geq 0 \; \forall i = 1, \ldots, K \text{ and } \sum_{i=1}^{K} c_i = 1 \right\}. \tag{24}$$

This yields the joint predictive distribution

$$p_t(r_{t+1}|\mathcal{D}) := p(r_{t+1}|R_t, \mathcal{H}_t, \mathcal{D}) = \int (c_t \otimes I)' p\left(r_{\mathcal{D},t+1}^{\text{vec}}|R_t, \mathcal{H}_t, \mathcal{D}\right) dr_{\mathcal{D},t+1}^{\text{vec}}, \tag{25}$$

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), decision-based model combinations (Billio et al., 2013) and optimal prediction pooling (Geweke and Amisano, 2011). In line with the latter, 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,

$$c_t^* = \arg\max_{c \in \Delta_{[0,1]^K}} \sum_{i=t-h_c}^{t} \log \left[\sum_{k=1}^{K} c_k p\left(r_i|R_{i-1}, \mathcal{H}_{i-1}, \mathcal{M}_k\right) \right], \tag{26}$$

where h_c is the window size and $p\left(r_i|R_{i-1}, \mathcal{H}_{i-1}, \mathcal{M}_k\right)$ is defined as equation in (22).[11] If the predictive density concentrates around the observed return values, the predictive likelihood is higher.[12] 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. We employ MCMC methods to generate draws from the posterior predictive distribution $r_{\mathcal{M}_k,t+1}^{(j)}$ and approximate the integral in optimization problem (EU) by means of Monte-Carlo techniques to compute an optimal portfolio $\hat{\omega}_{t+1}^{\mathcal{M}_k}$. A detailed description on the computational implementation of this approach is provided in Section A.5 of the Online Appendix.

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*

---

11. In our empirical application, we set $h_c = 250$ days.
12. Alternatively, we implemented utility-based model combination in the spirit of Billio et al. (2013) by choosing c_t^* as a function of past portfolio-performances net of transaction costs. However, the combination scheme resulted in very instable combination weights, putting mas on corner solutions. We therefore refrain from reporting results.

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).[13] Third, as a candidate representing the class of shrinkage estimators, we employ an approach based on Ledoit and Wolf (2003, 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 5.1. 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 volatilities and covariances.[14] 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. Stocks are separated into 4 equal-sized groups according to their average number of daily mid-quote observations. The resulting covariance matrix is then decomposed into $b = 10$ blocks representing pair-wise correlations within each group and across groups.[15] 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

$$r\tau_{t,1}^b := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,1}^{(i)} \right\}, \qquad r\tau_{t,l+1}^b := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,N^{(i)}(r\tau_{t,l}^b)}^{(i)} \right\}, \tag{27}$$

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

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

---

13. 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.

14. See, e.g., Andersen and Bollerslev (1998), Andersen et al. (2003) and Barndorff-Nielsen et al. (2009), among others.

15. 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.

which is observed under noise. Applying the multivariate realized kernel on each block of the covariance matrix, we obtain

$$K_t^b := \sum_{h=-L_t^b}^{L_t^b} k\left(\frac{h}{L_t^b + 1}\right) \Gamma_t^{h,b}, \tag{28}$$

where $k(\cdot)$ is the Parzen kernel, $\Gamma_t^{h,b}$ is the h-lag auto-covariance matrix of the assets log returns belonging to block $\mathcal{I}_b$, and L_t^b 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

$$\hat{H}_t^{\mathrm{b}} = \left(V_t^{\mathrm{b}}\right)^{-1} K_t^{\mathrm{b}} \left(V_t^{\mathrm{b}}\right)^{-1}, \qquad V_t^{\mathrm{b}} = \mathrm{diag}\left[K_t^b\right]^{1/2}. \tag{29}$$

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

$$\hat{\Sigma}_t^{\mathrm{BRK}} = \mathrm{diag}\left(\hat{\sigma}_{t,1}^2, \ldots, \hat{\sigma}_{t,N}^2\right)^{1/2} \hat{H}_t \mathrm{diag}\left(\hat{\sigma}_{t,1}^2, \ldots, \hat{\sigma}_{t,N}^2\right)^{1/2}. \tag{30}$$

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

We parametrize a suitable return distribution, which is driven by the dynamics of $\hat{\Sigma}_{S,t}^{BRK}$ and is close in the spirit to 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 an inverse Wishart distribution.[16] Thus, the model is defined by:

$$\mathcal{L}(r_{t+1}|\Sigma_{t+1}) \sim N(0, \Sigma_{t+1}), \tag{31}$$

$$\Sigma_{t+1}|\kappa, B_t \sim W_N^{-1}(\kappa, B_t), \tag{32}$$

$$\kappa B_t = \hat{\Sigma}_{S,t}^{\mathrm{BRK}}, \tag{33}$$

$$\kappa \sim \exp(100)\,\mathbb{1}_{\kappa > N-1}, \tag{34}$$

---

16. This represents a multivariate extension of the normal-inverse-gamma approach, applied, for instance, by Andersson (2001) and Forsberg and Bollerslev (2002).

where $W_N^{-1}(\kappa, B_t)$ denotes a N-dimensional inverse-Wishart distribution with degrees of freedom κ and scale matrix B_t, $\exp(\lambda)$ denotes an exponential distribution with rate $\lambda(= 100)$ and $\mathbb{1}_{\kappa > N-1}$ denotes a truncation at $N - 1$. Though 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.

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.[17] Likewise, stochastic volatility (SV) 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) propose means to reduce the enormous computational burden for high dimensional estimations of SV objects. 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

$$\tilde{r}_t = \Lambda V(\xi_t)^{1/2}\zeta_t + Q(\xi_t)^{1/2}\varepsilon_t, \tag{35}$$

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. The $j \times j$ diagonal matrix $V(\xi_t) = \text{diag}\left(\exp(\xi_{N+1,t}), \ldots, \exp(\xi_{N+j,t})\right)$ captures 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

$$\text{cov}\left(r_t | \xi_t\right) = \Sigma_t(\xi_t) = \Lambda V_t(\xi_t)\Lambda' + Q_t(\xi_t), \tag{36}$$

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

---

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

are triggered by j common factors. All $N + j$ latent factors are assumed to follow AR(1) processes,

$$\xi_{it} = \mu_i + \phi_i(\xi_{t-1,i} - \mu_i) + \sigma_i\eta_{t,i} \ i = 1, \ldots, N + j, \tag{37}$$

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 distributions 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, $\pi(\mu_i, \phi_i, \sigma_i)$ are in line with Aguilar and West (2000): 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 $\sigma_i^2 \sim G\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$.[18] For each element of the factor loadings matrix, a hierarchical zero-mean Gaussian distribution is chosen.

4.3 Covariance Shrinkage

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

$$S_t := \frac{1}{h-1} \sum_{i=t-h}^{t} (r_i - \hat{\mu}_t)(r_i - \hat{\mu}_t)', \tag{38}$$

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 (2003, 2004) propose shrinking S_t towards a more efficient (though biased) estimator of Σ_t.[19] The classical linear shrinkage estimator is given by

$$\hat{\Sigma}^{t,\text{Shrink}} = \hat{\delta}F_t + (1 - \hat{\delta})S_t, \tag{39}$$

where F_t denotes the sample constant correlation matrix and $\hat{\delta}$ minimizes the Frobenius norm between F_t and S_t. 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

---

18. 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). See Kastner (2018) for further details.
19. 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). This is left for future research.

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 $p_t\left(r_{t+1}|\hat{\Sigma}^{t,Shrink}\right) \sim N\left(0, \hat{\Sigma}^{t,Shrink}\right)$. Equivalently, a Gaussian framework is implemented for the sample 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 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 have been 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.[20] 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.[21] All the computations are performed after adjusting for stock splits and dividends. We extend our data set by HF-data extracted from the LOBSTER database[22], which provides tick-level message data for every asset and trading day. Utilizing midquotes amounts to more than 73 billion observations.

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. More detailed descriptions of the dataset and the computations are provided in Section A.6 of the Online Appendix. The distinct steps of the estimation and forecasting procedure are as follows: We implement $K = 4$ different models as of Section 4. The HF approach is based on the smoothed BRK-Wishart model with 4 groups, the SV model is based on $j = 3$ common factors, while forecasts based on the sample covariance matrix S_t and its regularized version $\hat{\Sigma}^{t,\,Shrink}$ are computed using a rolling window size of 500 days.[23]

Moreover, in line with a wide range of studies utilizing daily data, we refrain from predicting

---

20. 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.
21. Returns are computed as log-returns in our analysis. As we work with daily data, however, the difference to simple returns is negligible.
22. See https://lobsterdata.com.
23. The predictive accuracy of the SV model is very similar for values of j between 1 and 5, but declines when including more factors.

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. We sequentially generate (MCMC-based) samples of the predictive return distributions to compute optimal weights for every model. For the model combinations, we sequentially update the combination weights based on past predictive performance to generate the optimal allocation vector $\hat{\omega}_{t+1}^{\mathcal{D}}$.

In order to quantify the robustness and statistical significance of our results, we perform a bootstrap analysis by re-iterating the procedure described above 200 times for random subsets of $N = 250$ stocks out of the total 308 assets.[24]

5.2 Evaluation of the Predictive Performance

In a first step, we evaluate the predictive performance of the individual models. Note that this step does not require to compute any portfolio weights.[25] A popular metric to evaluate the performance of predictive distributions is the log posterior predictive likelihood $\log p(r_t^O | R_{t-1}, \mathcal{H}_{t-1}, \mathcal{M}_k)$, where r_t^O are the observed returns at day t, indicating how much probability mass the predictive distribution assigns to the observed outcomes. Table 1 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 covariance matrix solely is not sufficient to provide accurate forecasts. Shrinking the covariance matrix, 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 similarly to the HF model is an interesting finding as it utilizes only daily data and thus much less information than the latter. 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.[26]

---

24. 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 more than one month computing time on a strong cluster such as the Vienna Scientific Cluster.

25. A detailed visual analysis of the model's forecasts of the high-dimensional return distribution based on generated samples from the posterior predictive distribution is provided in Section A.7 of the Online Appendix.

26. 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. 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.

The last row of Table 1 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 (26) and evaluating the predictive density, $LS_{t+1}^{\text{Comb.}} := \log\left(\sum_{k=1}^{K} c_{t,k} p\left(r_{t+1}^O | R_t, \mathcal{H}_t, \mathcal{M}_k\right)\right)$, 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. Thus, our results confirm also in a high-dimensional setting the benefits of mixing high- and low-frequency data for financial time series, documented among others, by Ghysels et al. (2006), BańBura et al. (2013) and Halbleib and Voev (2016).

Figure 2 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 2 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 covariance matrix are negligible and are always dominated by the shrinkage estimator.

Superior predictive accuracy, however, is not equivalent to superior portfolio performance. An investor *not* adjusting for transaction costs ex ante may rebalance her portfolio unnecessarily often if she relies on quickly adjusting predictive models such as SV 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 is 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 compute portfolio performances based on bootstrapped portfolio weights for the underlying asset universe consisting 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 $\gamma = 4$. All computations are based on an investor with

power utility function $U_\gamma(r) = \frac{(1+r)^{1-\gamma}}{1-\gamma}$.[27]

After holding the assets for a day, she realizes the gains and losses, updates the posterior distribution and recomputes optimal portfolio weights. This procedure is repeated for each period and allows analyzing the time series of the realized ("out-of-sample") returns $r_{t+1}^k = \sum_{i=1}^{N} \hat{\omega}_{t+1,i}^{\mathcal{M}_k} r_{t+1,i}^O$. 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 (11). 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.

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

$$r_{t+1}^{k,\text{nTC}} = r_{t+1}^k - \nu_{L_1}\left(\hat{\omega}_{t+1}^{\mathcal{M}_k}\right) = \sum_{i=1}^{N} \hat{\omega}_{t+1,i}^{\mathcal{M}_k} r_{t+1,i}^O - \nu_{L_1}\left(\hat{\omega}_{t+1}^{\mathcal{M}_k}\right), \quad t = 1, \ldots, T. \tag{40}$$

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

$$\hat{\mu}\left(r^{k,\text{nTC}}\right) := \frac{1}{T} \sum_{t=1}^{T} r_t^{k,\text{nTC}}, \tag{41}$$

$$\hat{\sigma}\left(r^{k,\text{nTC}}\right) := \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(r_t^{k,\text{nTC}} - \hat{\mu}\left(r^{k,\text{nTC}}\right)\right)^2}, \tag{42}$$

$$SR^k := \frac{\hat{\mu}\left(r^{k,\text{nTC}}\right)}{\hat{\sigma}\left(r^{k,\text{nTC}}\right)}, \tag{43}$$

---

27. Whereas the optimization framework (EU) does not generally depend on the power utility function, Bayesian numerical methods (as MCMC) allow us to work with almost arbitrary utility functions. We feel, however, that the specific choice of a power utility function is not critical for the overall quality of our results: First, as advocated, among others, by Jondeau and Rockinger (2006) and Harvey et al. (2010), power utility implies that decision-making is unaffected by scale, as the class of iso-elastic utility function exhibits constant relative risk aversion (CRRA). Furthermore, in contrast to quadratic utility, power utility is affected by higher order moments of the return distribution, see also Holt and Laury (2002). Thus, power utility allows us to study the effect of parameter- and model uncertainty and makes our results more comparable to the existing literature.

$$CE^k := 100 \left(\left(\frac{1}{T} \sum_{t=1}^{T} \left(1 + r_t^{k,\text{nTC}} \right)^{1-4} \right)^{\frac{1}{1-4}} - 1 \right). \tag{44}$$

Moreover, we quantify the portfolio turnover, the average weight concentration, and the average size of the short positions:

$$TO^k := \frac{1}{T-1} \sum_{t=2}^{T} \left\| \hat{\omega}_t^{\mathcal{M}_k} - \frac{\hat{\omega}_{t-1}^{\mathcal{M}_k} \circ (1 + r_t)}{1 + \hat{\omega}_{t-1}^{\mathcal{M}_k'} r_t} \right\|_1, \tag{45}$$

$$pc^k := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\hat{\omega}_{t,i}^{\mathcal{M}_k} \right)^2, \tag{46}$$

$$sp^k := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} |\hat{\omega}_{t,i}| \mathbb{1}_{\left\{ \hat{\omega}_{t,i}^{\mathcal{M}_k} < 0 \right\}}. \tag{47}$$

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 and bi-monthly rebalancing.[28] We also include the "classical" global minimum variance portfolio based on the Ledoit-Wolf shrinkage estimator, $\omega_{t+1}^{\text{mvp}} := \frac{\left(\hat{\Sigma}^{\text{t, Shrink}} \right)^{-1} \iota}{\iota' \left(\hat{\Sigma}^{\text{t, Shrink}} \right)^{-1} \iota}$, and the optimal global minimum variance weights with a no-short sale constraint, computed as solution to the optimization problem:

$$\omega_{t+1}^{\text{mvp, no s.}} = \arg\min \omega' \hat{\Sigma}^{\text{t, Shrink}} \omega \text{ s.t. } \iota' \omega = 1 \text{ and } \omega_i \geq 0 \ \forall i = 1, \ldots, N. \tag{48}$$

Furthermore, we include portfolio weights ω_{t+1}^{ϑ} computed as the solution to the optimization problem (18) for gross-exposure constraints $\vartheta \in \{1, \ldots, 8\}$ as proposed by Fan et al. (2012).

Finally, we compare the results to optimal combinations of portfolios which have been proposed in literature: Here, the idea of shrinkage is applied *directly* to the portfolio weights by constructing portfolios which optimally (usually with respect to expected utility) balance between the estimated portfolio and a specified target portfolio. Candidates from this field are a mixture of the efficient portfolio and the naive allocation according to Tu and Zhou (2011), with allocations denoted by ω^{tz}, and mixtures of the efficient portfolio and the minimum variance allocation according to Kan and Zhou (2007), with allocations denoted by ω^{kz}. Out of strategies considering estimation risk of the parameters specifically, we utilize the approach of Jorion (1986). Here, shrinkage is applied to the estimated moments of the return distribution via traditional Bayesian estimation methods, with the resulting estimates used as inputs to compute the efficient portfolio (ω^{bs}). Finally, we create a buy-and-hold portfolio consisting of an investment in all 308 assets, weighted by the market capitalization as of March 2007. This benchmark strategy aims at replicating the market portfolio. Detailed descriptions of the implementation of the individual strategies are provided in Section A.8

---

28. Implementing naive schemes based on weekly, monthly or no rebalancing at all, does not qualitatively alter the results.

of the Online Appendix.

Table 2 summarizes the (annualized) results. The results in the first panel correspond to strategies ignoring transaction costs in the portfolio optimization by setting $\nu_{L_1}(\omega) = 0$ when computing optimal weights ω^*. These strategies employ the information conveyed by the predictive models, but ignore transaction costs. However, after rebalancing, 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 optimization process.

As indicated by highly negative average portfolio returns, a priori turnover penalizing is crucial in order to obtain a reasonable portfolio performance in a setup with transaction costs. The major reason is not necessarily the use of a sub-optimal weight vector, but rather the 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 rebalancing. All four predictive models generate average annualized portfolio turnover of more than 50%, 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. 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.[29]

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 competing 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 performance, 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. *Combining* forecasts, however, outperforms individual models and yields higher Sharpe ratios. 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.

---

29. 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.

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. These findings underline our conclusions drawn from Proposition 5: Though gross-exposure constraints are closely related to a penalization of transaction costs and minimize the effect of misspecified elements in the 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 2 gives the average fraction of days where the portfolio is rebalanced.[30] The incorporation of transaction costs implies that investors rebalance in just roughly 15% of all trading days. 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 rebalancing. In contrast, the benchmark strategies that do not account for transaction costs tend to rebalance permanently (small fractions of) wealth, thus cumulating an extraordinarily high amount of transaction costs.

To interpret the economic significance of the out-of-sample portfolio performance, 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 $\mathcal{M}_2$.[31] The fee is computed such that the investor would be indifferent between the two strategies in 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

$$\sum_{t=1}^{T} \left(1 + r_t^{\mathcal{M}_1, nTC}\right)^{(1-\gamma)} = \sum_{t=1}^{T} \left(1 + r_t^{\mathcal{M}_2, nTC} - \Delta_{\mathcal{M}_1, \mathcal{M}_2}\right)^{(1-\gamma)}. \tag{49}$$

Table 3 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. 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, the switching fees between the individual penalized models are substantially lower. This result again confirms the

---

30. Due to numerical instabilities, we interpret a turnover of less than 0.001% as no rebalancing.
31. We also used alternative values of γ with $\gamma = 1, \ldots, 10$, however, do not find qualitative differences.

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. In particular, our theoretical framework shows the close relation between Lasso- (Ridge-) penalization of turnover in the presence of proportional (quadratic) transaction costs. The implied turnover penalization improves portfolio allocations in terms of Sharpe ratios and utility gains. This is on the one hand due to regularization effects improving the stability and the conditioning of covariance matrix estimates and on the other hand due to a significant reduction of the amount (and frequency) of rebalancing 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 and employing more than 70 billion intra-daily trading message observations, 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. 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, a portfolio bootstrap reveals 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. 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. Finally, we find that strategies that *ex-ante* account for transaction costs significantly outperform a large range of well-known competing strategies in the presence of transaction costs and many assets.

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.

7 Acknowledgements

References

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

Akian, M., J. L. Menaldi, and A. Sulem (1996). On an investment-consumption model with transaction costs. *SIAM Journal on control and Optimization 34*(1), 329–364.

Andersen, T. G. and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review 39*(4), 885–905.

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

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.

Balduzzi, P. and A. W. Lynch (2000). Predictability and transaction costs: The impact on rebalancing rules and behavior. *The Journal of Finance 55*(5), 2285–2309.

BańBura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. *Handbook of Economic Forecasting 2*(Part A), 195–237.

Bao, Y., T.-H. Lee, and B. Saltoğlu (2007). Comparing density forecast models. *Journal of Forecasting 26*(3), 203–225.

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (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., P. R. Hansen, N. Shephard, and A. Lunde (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 N. Shephard (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 C. W. J. Granger (1969). The combination of forecasts. *Or 20*(4), 451–468.

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

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. Working paper.

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

Campbell, J. Y., Y. L. Chan, and L. M. Viceira (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics 67*(1), 41–80.

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

Chopra, V. K. and W. T. Ziemba (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 A. R. Norman (1990). Portfolio selection with transaction costs. *Mathematics of operations research 15*(4), 676–713.

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

DeMiguel, V., A. Martín-Utrera, and F. J. Nogales (2015). Parameter uncertainty in multiperiod portfolio optimization with transaction costs. *Journal of Financial and Quantitative Analysis 50*(6), 1443–1471.

DeMiguel, V., F. J. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies 27*(4), 1031–1073.

DeMiguel, V. and A. V. Olivares-Nadal (2018). Technical Note—A Robust Perspective on Transaction Costs in Portfolio Optimization. *Operations Research* (Forthcoming).

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

Engle, R. F., O. Ledoit, and M. Wolf (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 0*(0), 1–13.

Engle, R. F., F. Robert, and R. Jeffrey (2012). Measuring and modeling execution cost and risk. *Journal of Portfolio Management 38*(2), 14–28.

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

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

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

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

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

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130–141.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*(1-2), 59–95.

Halbleib, R. and V. Voev (2016). Forecasting Covariance Matrices: A Mixed Approach. *Journal of Financial Econometrics 14*(2), 383–417.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

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

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010). Portfolio selection with higher moments. *Quantitative Finance 10*(5), 469–485.

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

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

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science 14*(4), 382–401.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *The American Economic Review 92*(5), 1644–1655.

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

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

Jobson, J. D. and B. Korkie (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association 75*(371), 544–554.

Jondeau, E. and M. Rockinger (2006). Optimal portfolio allocation under higher moments. *European Financial Management 12*(1), 29–55.

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

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

Kastner, G. (2018). Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics* (Forthcoming).

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017). Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models. *Journal of Computational and Graphical Statistics 26*(4), 905–917.

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

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

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

Leland, H. (1999). Optimal portfolio management with transactions costs and capital gains taxes. Working paper.

Liu, H. (2004). Optimal consumption and investment with transaction costs and multiple risky assets. *The Journal of Finance 59*(1), 289–338.

Lynch, A. W. and S. Tan (2010). Multiple risky assets, transaction costs, and return predictability: Allocation rules and implications for us investors. *Journal of Financial and Quantitative Analysis 45*(4), 1015–1053.

Magill, M. J. P. and G. M. Constantinides (1976). Portfolio selection with transactions costs. *Journal of Economic Theory 13*(2), 245–263.

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

Pflug, G. C., A. Pichler, and D. Wozabal (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.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Taksar, M., M. J. Klass, and D. Assaf (1988). A diffusion model for optimal portfolio selection in the presence of brokerage fees. *Mathematics of operations research 13*(2), 277–294.

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

Tóth, B., Y. Lemperiere, C. Deremble, J. de Lataillade, J. Kockelkoren, and J.-P. Bouchaud (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X 1*(2), 021006.1–021006.11.

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

Uppal, R. and T. Wang (2003). Model Misspecification and Underdiversification. *The 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 S. Shi (2000). Predicting daily probability distributions of S&P500 returns. *Journal of Forecasting 19*(4), 375–392.

31

	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 1: Predictive accuracy of the models, evaluated using the time series of out-of-sample predictive log likelihoods, corresponding to $\log p_t(r_{t+1}^O | \mathcal{M}_k)$. 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 (26).

	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	TO	pc	sp	% Trade
$\omega^{\text{SV, no TC}}$	-58.9 [-62.7,-55.8]	16.12 [15.72,16.52]	-	-	55.73 [53.52,58.16]	31.102 [30.279,31.942]	2.41 [2.12,2.78]	> 99.95
$\omega^{\text{HF, no TC}}$	< -99	15.97 [15.52,16.33]	-	-	245.4 [239.29,250.43]	56.791 [55.7,57.892]	4.01 [3.83,4.65]	> 99.95
ω^{Sample}	5.4 [3.8,7.2]	16.50 [15.87,17.16]	0.326 [0.224,0.446]	3.993 [2.367,5.908]	0.19 [0.18,0.20]	42.427 [37.75,47.465]	1.21 [0.98,1.38]	18.45 [18.20,19.74]
ω^{LW}	5.6 [4.5,6.6]	14.72 [14.28,15.06]	0.380 [0.307,0.449]	4.501 [3.412,5.497]	0.88 [0.83,0.93]	24.405 [22.784,26.427]	0.98 [0.65,1.11]	11.95 [10.92,12.96]
ω^{HF}	6.5 [5.9,6.9]	16.71 [16.52,16.93]	0.388 [0.354,0.414]	5.081 [4.488,5.518]	0.14 [0.09,0.21]	17.165 [15.937,18.062]	0.48 [0.21,0.61]	14.44 [13.41,15.54]
ω^{SV}	6.6 [5.9,7.0]	16.76 [16.59,16.99]	0.391 [0.354,0.417]	5.153 [4.489,5.589]	0.32 [0.21,0.41]	17.860 [16.086,18.827]	0.71 [0.68,0.92]	14.13 [13.11,15.20]
$\omega^{\text{Comb.}}$	6.5 [5.0,7.8]	14.86 [14.45,15.39]	0.438 [0.342,0.52]	5.401 [3.877,6.693]	0.49 [0.46,0.53]	20.609 [18.483,23.115]	0.81 [0.54,0.92]	13.70 [12.53,14.83]
ω^{Naive}	5.2 [4.9,5.5]	23.5 [23.14,23.84]	0.224 [0.204,0.234]	2.399 [2.035,2.733]	1.10 [1.09,1.12]	0.4 [0.4,0.4]	-	> 99.95
$\omega^{\text{Naive, 2m}}$	5.9 [5.7,6.2]	23.03 [22.68,23.36]	0.258 [0.245,0.273]	3.294 [2.991,3.632]	0.15 [0.14,0.16]	0.403 [0.403,0.404]	-	1.66 [1.63,1.66]
ω^{mvp}	-59.4 [-62.8,-56.1]	14.11 [13.73,14.55]	-	-	53.49 [52.25,54.64]	45.65 [42.864,48.28]	3.33 [3.24, 3.42]	> 99.95
$\omega^{\text{mvp, no s.}}$	2.8 [2.0,3.6]	13.1 [12.97, 13.29]	0.213 [0.149, 0.279]	1.935 [1.092, 2.794]	3.53 [3.34, 3.69]	10.616 [9.161, 11.548]	-	> 99.95
$\omega^{\text{kz, 3f}}$	< -99	> 50	-	-	941.9 [175.51,2339.7]	> 1000	5.98 [3.72, 8.8]	> 99.95
ω^{tz}	< -99	> 50	-	-	> 1000	> 1000	28.09 [17.31, 57.51]	> 99.95
ω^{bs}	< -99	> 50	-	-	> 1000	> 1000	124.7 [39.8, 228.14]	> 99.95
ω^{ϑ}	-24.28 [-26.28, -22.15]	12.04 [11.78, 12.34]	-	-	25.846 [25.43, 26.31]	18.692 [18.05, 19.28]	1.499 [1.40, 1.58]	> 99.95
ω^{Market}	6.2	21.19	0.292	3.958	-	-	-	-

Table 2: 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 (11)). $\hat{\mu}$ is the annualized portfolio return in percent, $\hat{\sigma}$ is 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 weight 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%.

	Naive	LW	Sample	SV	HF	Combination
mvp no s.	1.98 [1.86,2.01]	6.37 [6.24,7.51]	6.16 [5.98,7.56]	6.01 [5.80,6.89]	6.41 [6.32,7.25]	7.73 [7.12,8.02]
Naive		6.18 [5.65,7.86]	5.99 [5.41,6.43]	5.83 [5.44,6.46]	6.40 [6.29,6.82]	7.57 [7.25,7.89]
LW			0.01 [-0.29,0.60]	0.41 [-0.01,0.91]	0.11 [-0.43,0.86]	1.44 [0.57,2.25]
Sample				0.1 [-0.22,0.89]	0.3 [-0.12, 1.32]	1.38 [0.84,2.13]
SV					- 0.68 [-0.29, 1.02]	1.11 [0.95, 1.81]
HF						1.31 [0.88,1.79]

Table 3: 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 7.57 basis point to switch from the naive portfolio to the combination strategy. The number in brackets denote the bootstrapped 95 % confidence intervals.*
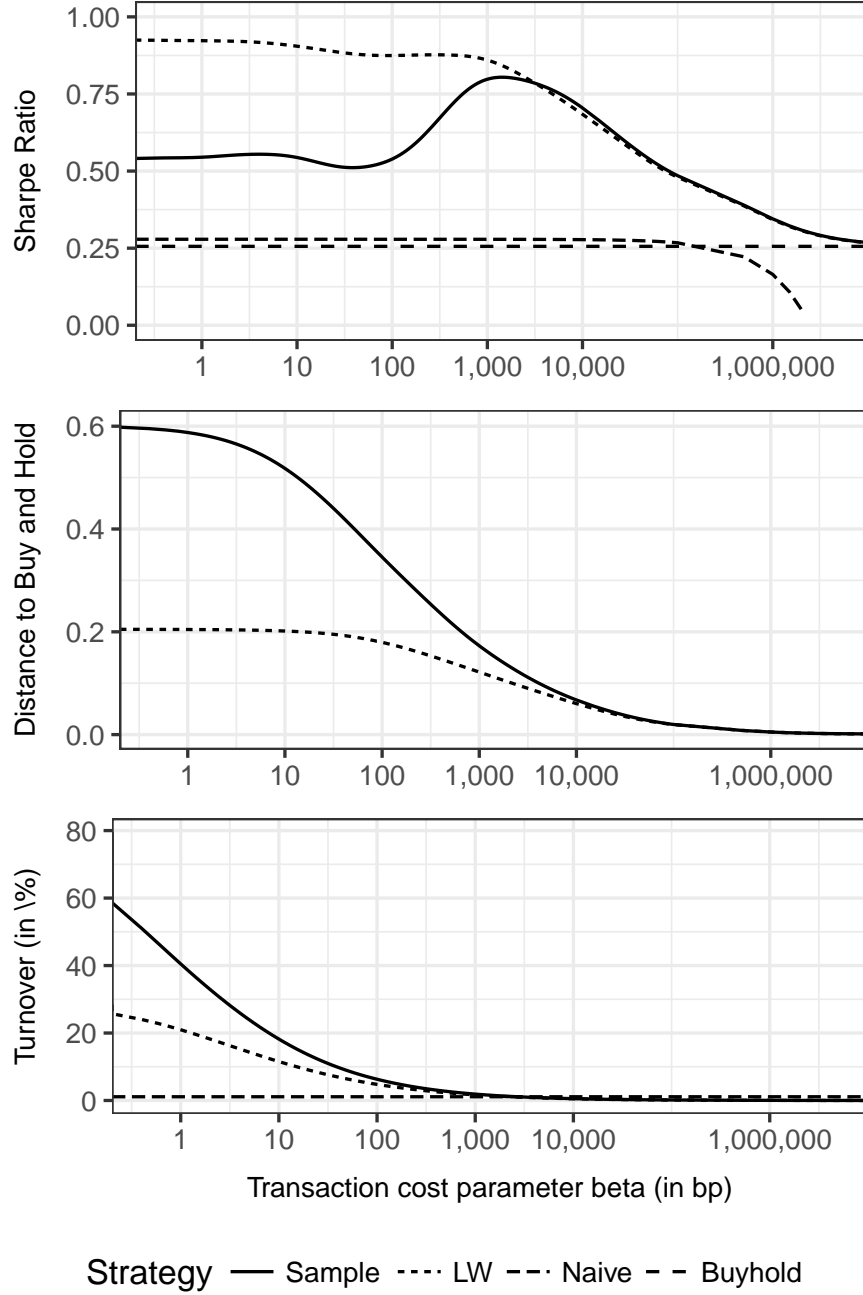
Figure 1: Annualized realized Sharpe ratio net of transaction costs for portfolios based on $N = 308$ assets with daily reallocation for the period from June 2007 until March 2017 with quadratic transaction costs. The portfolio weights are computed as solutions to optimization problem (3) with varying choices of the transaction cost parameter β (in basis points). The covariance matrix is estimated using the sample covariance matrix S_t and a regularized version $\hat{\Sigma}^{t,\text{Shrink}}$ (Ledoit and Wolf, 2003, 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 to the buy-and-hold portfolio with naive $(1/N)$ initial allocation. The third panel illustrates average turnover (in percent), measured as L_1 difference between ω_{t+1}^β and $\omega_{t^+}^\beta$. Strategy *Naive* denotes an allocation with daily rebalancing to the naive allocation, *Buyhold* corresponds to implementing $1/N$ once and refraining from rebalancing afterwards.
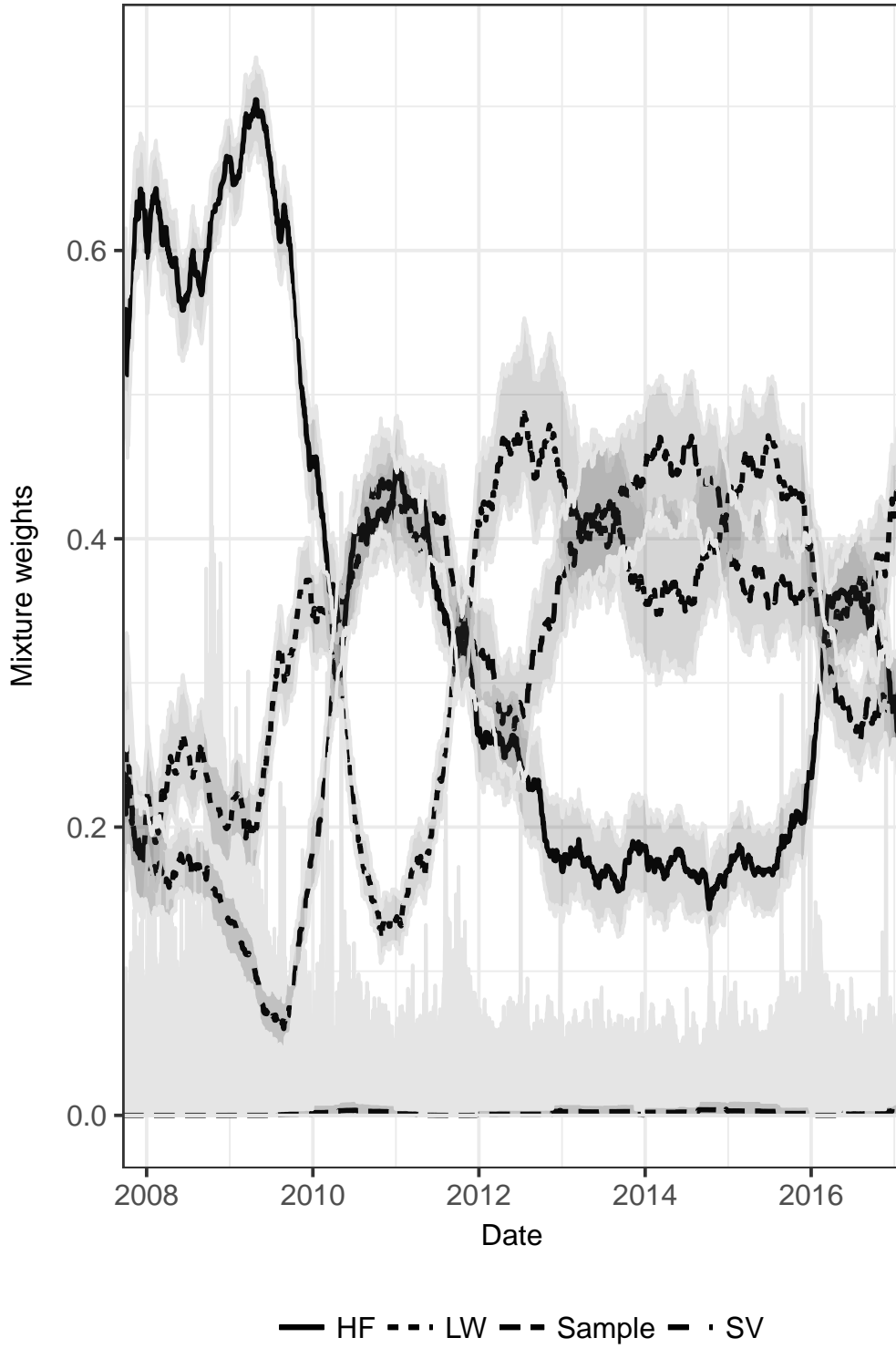
Figure 2: Combination weights computed based on (26). The weights are sequentially updated to maximize the log posterior predictive likelihood of past observed returns based on a rolling-window of 250 days. The lines correspond to the mean of the bootstrapped combination weights. The gray confidence 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.