

Hautsch, Nikolaus; Voigt, Stefan

**Conference Paper**

## Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty: Adaptive Mixing of High- and Low-Frequency Information

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Financial Econometrics I, No. F14-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Hautsch, Nikolaus; Voigt, Stefan (2017) : Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty: Adaptive Mixing of High- and Low-Frequency Information, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Financial Econometrics I, No. F14-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168222>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

---

# Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty: Adaptive Mixing of High- and Low-Frequency Information

Nikolaus Hautsch,<sup>\*</sup> Stefan Voigt<sup>†</sup>

---

**Preliminary version. Please do not circulate.**

We propose a Bayesian sequential learning framework for high-dimensional asset allocations under model ambiguity and parameter uncertainty. We consider portfolio allocations maximizing predictive expected utility after transaction costs, optimally balancing implementation shortfall and adjustments due to updated information. The unifying framework allows for time-varying mixtures of predictive return distributions which may exhibit fat tails, resulting from high- and low-frequency data. The model is estimated via MCMC methods and allows for a wide range of data sources as inputs. We consider predictive models resulting from high-dimensional Wishart approaches for high-frequency based blocked realized kernels, low-frequency based multivariate stochastic volatility factor models and regularized daily covariance estimates. Employing the proposed framework on a large set of NASDAQ-listed stocks, we observe that time-varying mixtures of high- and low-frequency based return predictions significantly improve the out-of-sample portfolio performance compared to individual models and outperform the naive 1/N-allocation in terms of Sharpe ratio and utility-based measures. Bootstrapping the optimization procedure shows that our results are robust with respect to the choice of the asset universe. We show that regularization of turnover is crucial in large dimensions and illustrate that the relative contribution of high-frequency data and low-frequency data strongly varies over time.

**JEL Classification:** C52, C11, C58, G11

**Keywords:** Portfolio choice, model uncertainty, estimation uncertainty, transaction costs, model combination, high frequency data

---

<sup>\*</sup>University of Vienna, Department of Statistics and Operations Research.

<sup>†</sup>WU (Vienna University of Economics & Business), Vienna Graduate School of Finance.

# 1. INTRODUCTION

Structuring a decision process for portfolio optimization in the spirit of Markowitz (1952) remains a challenge for econometricians and practitioners. A quantification of potential risk and rewards of future returns is not readily observable, instead, the investor can only make assumptions about the (possibly time-varying) data generating process (DGP) of future returns. Even if we could abstract from the question '*which parameters are driving the returns of tomorrow?*' it would still be necessary to answer '*what are the parameters?*'. An investor ignoring the effect of parameter- and model uncertainty on the predictions of future returns underestimates the risks she is exposed to (see e.g. Uppal and Wang (2003)). Good forecasts solely are not sufficient to form successful investment decisions. Transaction costs could prevent the portfolio adjustment because the expected benefits of reallocating wealth may be smaller than the costs associated with turnover.<sup>1</sup>

In this paper we show how to disentangle two driving forces behind promising portfolio allocation decisions, namely (i) improving predictive accuracy by adjusting for estimation uncertainty and (ii) balancing responsiveness to updated (high-frequency based) predictions and transaction costs.

Exploiting high-frequency data (HF) opens a channel to increase the precision of covariance estimates and forecasts.<sup>2</sup> Predictions of return covariances exploiting HF data immediately react to changing market circumstances and are considered useful especially during times of abrupt changes in the volatilities and covariances (Hansen et al., 2012). Methods and HF data available to forecast covariances in large dimensions also triggered empirical studies putting emphasis on examining the benefits for portfolio selection, e.g. Fleming et al. (2003), Liu (2009) and Hautsch et al. (2015).<sup>3</sup> Their findings underline the benefits of intra-daily information to forecast covariances.

Asset allocations in practice, however, do not only require (precise) predictions of covariance matrices but predictive distributions beyond second moments. They moreover need to account for transaction costs, model ambiguity and parameter uncertainty. Moreover, changing market

---

<sup>1</sup>See Brandt (2010) for an excellent review on common pitfalls in portfolio optimization. The effect of ignoring estimation uncertainty is considered, amongst others, by Jobson et al. (1979), Jorion (1986), Chopra and Ziemba (1993) and DeMiguel et al. (2009). Model uncertainty is investigated by Wang (2005) and Garlappi et al. (2007).

<sup>2</sup>The benefits of high-frequency data to investigate covariances have been documented by a large range of applications, among others by Andersen and Bollerslev (1998), Andersen et al. (2001) and Barndorff-Nielsen (2002), Barndorff-Nielsen and Shephard (2004).

<sup>3</sup>Multivariate approaches are proposed, among others, by Barndorff-Nielsen et al. (2011), Bannouh et al. (2012), Fan et al. (2012), Noureldin et al. (2012), Aït-Sahalia and Xiu (2015), Liu et al. (2016) and Halbleib and Voev (2016).

regimes and structural breaks are hardly captured by a single model. Investors rather consider a set of predictive models and adjust their relative importance according to market conditions and past performance.

The objective of this paper is to analyze whether and to which extent adaptive combinations of different models are beneficial in high-dimensional portfolio problems under realistic conditions. We moreover aim studying the relative importance of high-frequency and low-frequency data in such problems and to which extent these contributions change over time. For this purpose, we propose a large-scale Bayesian sequential learning framework focusing on the entire return distribution instead of only the second moment (as in classical Markowitz global minimum variance setups). In addition, we consider transaction costs and parameter- and model-uncertainty in the decision process. Our approach ensures optimality of the portfolio weights with respect to the predicted out-of-sample utility after transaction costs and regularizes turnover conditional on current beliefs about the future evolution of the DGP of the returns.

We embed HF data as one potential source of information and combine the predictions with estimation methods based on lower frequencies. Mixing approaches considered are Bayesian Model Averaging (BMA) and optimal prediction pooling (Geweke and Amisano, 2011). These methods adapt to changes in the market environment based on the past predictive performance of the models. We also allow beliefs regarding the reliability of distinct models to be driven by an utility-based objective function in the spirit of Billio et al. (2013).

To our best knowledge, this paper is the first approach combining the benefits of different predictive models based on high-frequency and low-frequency data sources in a large-scale portfolio allocation framework under widely realistic conditions. Our approach brings together concepts from (i) Bayesian estimation for portfolio optimization, regularization and turnover penalization, (iii) predictive model combinations in high dimensions and (iv) high-frequency-based covariance modeling and prediction.<sup>4</sup>

In particular, we address the following research questions: **(i)** Can sequential learning and mixing of individual allocation strategies improve the out-of-sample portfolio performance in a high dimensional asset space? **(ii)** How important is turnover penalization in light of transaction costs?

---

<sup>4</sup>Bayesian estimation for portfolio optimization has been applied within a wide range of applications, starting with Brown (1976) and Jorion (1986). Imposing turnover penalties is related to the ideas of Brodie et al. (2009) and Gârleanu and Pedersen (2013). Tu and Zhou (2010), Tu and Zhou (2011) and Anderson and Cheng (2016) emphasize the benefits of model combination in portfolio decision theory. Sequential learning in a two-dimensional asset horizon is performed by Johannes et al. (2014). However, none of these approaches is focusing on mixtures of HF and lower frequencies approaches and aims at large dimensional allocation problems.

(iii) How strong is the time-variation in model combinations? (iv) What is the value of high-frequency predictions relative to low-frequency based predictions in a setting where not only second moments but the entire return distribution needs to be predicted?

We address these questions in a large-scale setting based on stocks traded on NASDAQ between 2007 and 2016. Utilizing high-frequency message data for the entire period, we compute covariance estimates using blocked realized kernels according to Hautsch et al. (2012). Precise (HF-based) covariance predictions solely, however, are shown to be insufficient when it comes to portfolio applications where the entire predictive density is needed. We therefore propose a Normal-Wishart model for block realized kernel covariances. Carefully adjusting for parameter uncertainty in the spirit of Forsberg and Bollerslev (2002) is shown to considerably improve the predictive accuracy of HF data-based estimates. In addition to HF data, we utilize predictive distributions resulting from low-frequency multivariate stochastic volatility factor models, capturing co-movement among assets and persistence in variances.<sup>5</sup>

We find that time-varying combinations of low-frequency-based and high-frequency-based predictive return distributions significantly increases the predictive performance compared to the use of all single models. We moreover show that the predictive superiority also translates into superior asset allocations. Based on out-of-sample portfolio returns and bootstrap inference for Sharpe ratios our approach outperform all implemented competitors. We also find that regularization of turnover is crucial in large dimensions in order to transform accurate predictions into high out-of-sample utilities: without considering potential transaction costs ex ante in the decision process it is not possible to obtain a satisfying performance. Finally, we find strong evidence for time-varying model weights. The relative contribution of high-frequency based return predictions is on average approximately 40% while the remaining 60% are provided by both the multivariate stochastic volatility model (approximately 30%) and predictions based on regularized sample covariances.

The structure of this paper is as follows: Section 2 introduces our baseline model to form coherent allocation decisions. Section 3 describes the combination framework to tackle model uncertainty in detail. Section 4 discusses high-frequency and low-frequency based predictive return distributions. Section 5 describes the data. Section 6 contains empirical results and Section 7 concludes.

---

<sup>5</sup>Stochastic volatility models have been shown to be beneficial in portfolio allocation by Aguilar and West (2000) and Han (2006).

## 2. DECISION FRAMEWORK AND ESTIMATION TECHNIQUES

We consider an investor who allocates her wealth among  $N$  distinct assets in the spirit of Markowitz (1952). At each point in time  $t$ , the investor observes past returns  $R_t = (r'_1, \dots, r'_t)' \in \mathbb{R}^{t \times N}$ , where  $r_i \in \mathbb{R}^N, i \in \{1, \dots, t\}$ . Additionally, the set of observable variables contains information  $\mathcal{F}_t$ , e.g. intra-daily message data or asset characteristics such as book-to-market ratios. Conditional on observing  $(R_t, \mathcal{F}_t)$  the investor draws inference about unobservable parameters of the return distribution with the aim to forecast the returns of tomorrow. Forecasts take the form of a multivariate probability distribution  $p_t(r_{t+1}|\mathcal{M}) := p(r_{t+1}|\mathcal{M}, R_t, \mathcal{F}_t)$  and express the beliefs about the DGP.  $\mathcal{M}$  captures the assumptions regarding the structure of the return distribution.

We define an optimal portfolio as an allocation which maximizes the expected utility of the investor after subtracting transaction costs arising due to rebalancing. Therefore, at time  $t$  the investor with utility function  $U_\gamma$  and risk aversion parameter  $\gamma$  solves a static maximization problem conditional on her current beliefs regarding the distribution of the returns of the next period:

$$\begin{aligned} \omega_{t+1}^* &:= \arg\max_{\omega \in \mathbb{R}^N, l'\omega=1} E(U_\gamma(\omega' r_{t+1} - v_t(\omega)) | \mathcal{M}, R_t, \mathcal{F}_t) & (\text{EU}) \\ &= \arg\max_{\omega \in \mathbb{R}^N, l'\omega=1} \int_{\mathbb{R}^N} U_\gamma(\omega' r_{t+1} - v_t(\omega)) p_t(r_{t+1}|\mathcal{M}) dr_{t+1}, & (1) \end{aligned}$$

where  $v_t(\omega)$  reflects transaction costs including broker fees or implementation shortfall. Rebalancing portfolio weights without taking into account a measure for transaction costs ignores that utility is derived after paying the broker and is not just based on the portfolio return  $\omega' r_{t+1}$ .<sup>6</sup>

One way to parametrize transaction costs is to model them as proportional to the amount of rebalancing  $v_t(\omega_{t+1}) \propto \|\omega_{t+1} - \omega_{t^+}\|$ . Here,  $\|\omega_{t+1} - \omega_{t^+}\|$  denotes some distance measure between the new allocation  $\omega_{t+1}$  and the allocation right before readjustment,  $\omega_{t^+} := \frac{\omega_t \circ r_t}{l'(\omega_t \circ r_t)}$ . The distinction between  $\omega_{t^+}$  and  $\omega_t$  is important, as the price dynamics in the interval  $[t, t+1]$  affect the allocation of wealth.

Penalizing transaction costs in such a way serves as natural shrinkage towards  $\omega_{t^+}$  and smooths portfolio weights. Shrinkage methods are commonly used to reduce prediction errors, whereas in our case deviating from  $\omega_{t^+}$  is beneficial if the updated beliefs regarding the parameters  $\Theta$  of the return distribution justify the potential costs. If a risk-free asset  $r_{f,t}$  is present, the optimization

---

<sup>6</sup>Whereas it is often the case that transaction costs are not directly considered within the decision framework, Perold (1984), Arnott and Wagner (1990) and Yoshimoto (1996) implement approaches based on the mean-variance framework by either constraining turnover (which may not result in an optimal allocation) or by adjusting expected portfolio returns for transaction costs.

approach can alternatively be rewritten as

$$\omega_{t+1}^* := \arg \max_{\omega \in \mathbb{R}^N} \int_{\mathbb{R}^N} U_\gamma(\omega' r_{t+1} + (1 - l' \omega) r_{f,t+1} - v_t(\omega)) p_t(r_{t+1} | \mathcal{M}) dr_{t+1}. \quad (2)$$

This notation does not require an explicit constraint as the  $N + 1$  weights are implicitly assured to sum-up to 1.

Under the very simplistic assumption of  $p_t(r_{t+1} | \mathcal{M})$  being multivariate normally distributed and ignoring  $v_t(\omega)$ , the optimization framework EU goes hand-in-hand with the initial Markowitz (1952) approach and allows to solve for  $\omega_{t+1}^*$  analytically.

## 2.1. THE NAIVE CASE: GAUSSIAN RETURNS WITH QUADRATIC TRANSACTION COSTS

The following illustrative example shows the effect of adjusting for expected turnover costs in the case when no further frictions are present. If we assume multivariate log-normal distributed returns with known parameters  $\Sigma$  and  $\mu$  and power utility function  $U_\gamma(r) = \frac{r^{1-\gamma}}{1-\gamma}$  the maximization problem simplifies a lot. We model the transaction costs  $v_t(\omega_{t+1}, \omega_{t^+}, \beta)$  for shifting wealth from  $\omega_{t^+}$  to allocation  $\omega_{t+1}$  as

$$v_t(\omega_{t+1}, \omega_{t^+}, \beta) = \beta(\omega_{t+1} - \omega_{t^+})'(\omega_{t+1} - \omega_{t^+}) \quad (3)$$

with some parameter  $\beta > 0$ . Therefore, at  $t + 1$  the investors' expectation of the parameters of the portfolio return is

$$\mu_{\text{PF}} = \omega'_{t+1} \mu - v_t(\omega_{t+1}, \omega_{t^+}, \beta) \text{ and } \sigma_{\text{PF}}^2 = \omega'_{t+1} \Sigma \omega_{t+1}. \quad (4)$$

The allocation  $\omega_{t+1}^*$  is chosen to maximize the Certainty Equivalent (CE) after transaction costs

$$\omega_{t+1}^* = \arg \max_{\omega \in \mathbb{R}^N, l' \omega = 1} \omega' \mu - v_t(\omega, \omega_{t^+}, \beta) - \frac{\gamma}{2} \omega' \Sigma \omega. \quad (5)$$

**Proposition 1.** *Optimal rebalancing  $\Delta^* = \omega_{t+1}^* - \omega_{t^+}$  based on optimization problem (5) is given by*

$$\Delta_\beta^* = \frac{1}{\gamma} \left( A^{-1} - \frac{1}{l' A^{-1} l} A^{-1} u' A^{-1} \right) \mu - A^{-1} \Sigma \omega_{t^+} + \frac{1}{l' A^{-1} l} A^{-1} u' A^{-1} \Sigma \omega_{t^+} \quad (6)$$

with  $A := \left( \frac{2}{\gamma} \beta I + \Sigma \right)$ .

*Proof.* See Appendix. □

$\Delta_\beta^*$  illustrates the trade-off between responding to available information (estimates of  $\mu$  and  $\Sigma$ ) and the costs resulting from constructing the corresponding efficient portfolio. The framework above reduces to the efficient portfolio if we ignore transaction costs (by choosing  $\beta = 0$ )

$$\omega_{t+1,\beta=0}^* = \Delta_{\beta=0}^* + \omega_{t^+} = \frac{1}{\gamma} \left( \Sigma^{-1} - \frac{1}{t' \Sigma^{-1} t} \Sigma^{-1} t t' \Sigma^{-1} \right) \mu + \underbrace{\frac{1}{t' \Sigma^{-1} t} \Sigma^{-1} t}_{\text{Global Min. Var. Portfolio}}. \quad (7)$$

Computing portfolio weights using  $\omega_{t+1,\beta=0}^*$  has been shown to perform rather poorly especially in a high dimensional asset space when substituting  $\mu$  and  $\Sigma$  with sample estimates  $\hat{\mu}$  and  $\hat{\Sigma}$ . Adjusting for transaction costs by choosing  $\omega_{t+1,\beta}^*$  reduces turnover and retains the information given by the return distribution parameters  $\hat{\mu}$  and  $\hat{\Sigma}$ . The turnover penalty can increase out-of-sample performance significantly and yield satisfying portfolio returns even if parameter uncertainty is not adjusted for and the sample estimates  $\hat{\mu}$  and  $\hat{\Sigma}$  do not reflect appropriate estimates. To illustrate this effect, Table 1 contains the results of a horse-race based on  $N = 100$  assets and daily readjustment, whereas we compute the portfolio weights by choosing  $\omega_{t+1,\beta}^*$  with  $\beta = 50/10.000$ .  $\hat{\Sigma}$  is computed as the sample variance-covariance estimator, a regularization approach (Ledoit and Wolf, 2004) and a stochastic volatility factor model.<sup>7</sup> Apparently, the allocation approach is able to allocate wealth in a manner which is satisfying: although frequently noted that the sample variance-covariance matrix is not reliable for the mean-variance approach (especially not for such a large asset horizon), turnover penalization restricts the transaction in a way which generates a positive Sharpe-Ratio after adjusting for transaction costs. In addition, as soon as we elaborate more sophisticated estimates such as the regularization approach proposed by Ledoit and Wolf (2004), we are able to beat the naive 1/N-portfolio. If we, however, relax the assumption of quadratic transaction costs, we cannot provide analytic solutions for the optimal portfolio. Furthermore, adjusting for model uncertainty requires deviating from simple CE maximization. Instead, we aim at considering the full predictive return distribution, also accounting for higher moments of the return distribution.

## 2.2. A STEP TOWARDS REALITY: INCORPORATING PARAMETER UNCERTAINTY

A crucial input to obtain sensible portfolio weights is the choice of the forecasting density of returns. Before taking into consideration model uncertainty, we focus on  $p_t(r_{t+1}|\mathcal{M})$  denoting the predictive return distribution of an investor who updates her beliefs sequentially according to Bayes theorem.

---

<sup>7</sup>A description of the dataset and the estimators is given in more detail in Section 6.



	$\Delta_{\beta=0}^*$			$\Delta_{\beta}^*$		
	Mean	SD	SR	Mean	SD	SR
Sample Covariance ( $\Delta_{\beta}^*$ )	-86.4	11.8	-	7.0	10.63	0.658
Ledoit Wolf ( $\Delta_{\beta}^*$ )	-67.3	10.8	-	8.3	11.41	0.727
Stochastic Volatility ( $\Delta_{\beta}^*$ )	-42.6	12.1	-	10.5	11.82	<b>0.888</b>
Naive				7.2	10.58	0.680

Table 1: Results of the empirical horse-race after out-of-sample after transaction costs based on 1904 trading days with imposed transaction fees of  $v(\omega_{t+1}) = 50/1000 \sum_{i=1}^N (\omega_{t+1,i} - \omega_{t^+,i})^2$ . *SR* denotes the (annualized) out-of-sample Sharpe ratio of the individual strategies. The values are annualized.

$\mathcal{M}$  contains a likelihood function of the returns  $\mathcal{L}(r_t|\Theta)$  with parameters  $\Theta$  and a proper choice of priors  $\pi(\Theta)$ . The stochastic nature of the unobservable parameters  $\Theta$  reflects that the investor does not know  $\Theta$  with certainty. Conditional on the set of available information,  $(R_t, \mathcal{F}_t)$ , learning with respect to the parameters is represented by the posterior distribution  $\pi(\Theta|R_t, \mathcal{F}_t) \propto \mathcal{L}(R_t|\Theta, \mathcal{F}_t)\pi(\Theta)$ . The distribution  $\pi(\Theta|R_t, \mathcal{F}_t)$  reflects the updated beliefs of the investor regarding the underlying parameters of the model after combining the inference drawn from the observed data with initial beliefs.

This forms the foundation to evaluate the (posterior) predictive distribution of the returns

$$p(r_{t+1}|\mathcal{M}, R_t, \mathcal{F}_t) := \int \mathcal{L}(r_{t+1}|\Theta)\pi(\Theta|R_t, \mathcal{F}_t)d\Theta. \quad (8)$$

Marginalizing out the uncertain parameters generates a fat-tailed distribution. Power utility functions take into account higher moments of the return distribution and therefore react sensitive to this specification. After observing the past performance of the assets and updating the beliefs about the parameter of the model, the investor chooses the optimal allocation of wealth by maximizing expected one-period ahead utility.

In general, maximization problem (EU) cannot be solved analytically. Instead we apply Monte Carlo methods to approximate the integral in the maximization problem (EU). This is done by computing  $M$  sample draws  $\theta^{(1)}, \dots, \theta^{(M)}$  from the posterior distribution  $\pi(\Theta|R_t, \mathcal{F}_t)$  via Markov Chain Monte Carlo algorithms.<sup>8</sup> Subsequently,  $M$  draws from the predictive distribution are generated by sampling  $r_{t+1}^{(m)}$  from  $\mathcal{L}(r_{t+1}|\mathcal{M}, \theta^{(m)})$ . Conditional on  $\omega \in \mathbb{R}^N$ , the integral in Equation

<sup>8</sup>See Hastings (1970) and Chib and Greenberg (1995).

(EU) can be approximated by

$$E(U_\gamma(\omega' r_{t+1} - v_t(\omega)) | \mathcal{M}, R_t, \mathcal{F}_t) \approx \frac{1}{M} \sum_{m=1}^M U_\gamma(\omega' r_{t+1}^{(m)} - v_t(\omega)). \quad (9)$$

The numerical solution  $\hat{\omega}_{t+1}$  of Equation (EU) is obtained by choosing  $\omega$  to maximize the sum in Equation (9). This approach is similar to the applications presented by Greyserman et al. (2006) and Han (2006).

Bayesian computational methods allow to draw inference about the distribution of non-linear transformations of our variable of interest. After generating draws from the posterior predictive distribution of the returns, we can directly infer the vector  $\{\hat{\omega}'_{t+1} r_{t+1}^{(m)} - v_t(\hat{\omega}_{t+1})\}_{m=1, \dots, M}$  which represents a draw from the posterior predictive portfolio return distribution (after accounting for transaction costs) conditional on model  $\mathcal{M}$ . Choosing  $\hat{\omega}_{t+1}$  by maximizing the predicted out-of-sample performance takes into account all the available information at time point  $t$  into account and penalizes large rebalancing by the choice of the function  $v_t(\omega)$ .

### 2.3. EXAMPLES

The following examples cover some of the most prominent classes of allocation strategies and show that many approaches can be reformulated as a solution to the general maximization problem (EU):

*Mean-variance approaches:* including the seminal work of Markowitz (1952), many approaches estimate portfolio weights based on the first two moments of the return distribution. This can be justified either by the assumption of multivariate Gaussian returns  $p_t(r_{t+1}) = N(\mu_t, \Sigma_t)$  or by a quadratic utility function. Proposed estimators for  $\Sigma_t$  include, among others, shrinkage approaches (Ledoit and Wolf, 2004). Neglecting higher moments of the return distribution, however, may deteriorate the performance of portfolio allocation strategies (see e.g. Harvey et al. (2010); Jondeau and Rockinger (2012)).

*Subjective approach of Black and Litterman:* Black and Litterman (1992) propose to form portfolio weights by taking into account two sources of information about future excess returns: the imposed market equilibrium based on CAPM and subjective views, which are both associated with uncertainty. They implicitly apply a Bayesian approach, whereas the distributional parameters are obtained via reverse optimization (see Idzorek (2002)).

The mean of the predictive returns,  $E(r_{t+1})$ , is assumed to be distributed with covariance proportional to the sample covariance  $\tau \hat{\Sigma}$ . Basic foundation is a Gaussian likelihood  $r_{t+1} \sim N(\mu_t, \hat{\Sigma}_t)$ .

The investor expresses her subjective views about  $\mu_t$  by imposing  $P\mu_t \sim N(q, \Omega_t)$  where  $\Omega$  is a diagonal matrix. The Black and Litterman approach rests on the assumption that under a complete non-informative prior ( $|\Omega| \rightarrow \infty$ ) the investor should stick to the market equilibrium portfolio. Using the market equilibrium models as a prior is equivalent to setting  $\mu_{\text{equ}} = \gamma \hat{\Sigma}_t \omega_{\text{mkt}}$ . Imposing a natural conjugate prior for  $\mu_t \sim N(\mu_{\text{equ}}, \lambda \hat{\Sigma}_t)$  leads to a mixed estimation model with corresponding updated posterior of  $\mu_t \sim N(m_v, \Lambda_v)$  and the Bayesian predictive density of one period ahead returns is obtained as  $r_{t+1}|v, \hat{\Sigma}_t \sim N(m_v, \hat{\Sigma}_t + \Lambda_v)$ . Transaction costs are not considered in the Black Litterman approach.

*Bayesian approaches:* Greyserman et al. (2006) propose to compute portfolio weights based on Equation (EU) within a fully Bayesian setting. Hereby, they rely on a hierarchical Gaussian framework for the returns. Within their empirical study, Greyserman et al. (2006) do not penalize transaction costs.

Han (2006) embeds the allocation framework into time-varying stochastic volatility return models, finding that neglecting stylized facts such as volatility clustering and time-varying volatility diminish predictive performance of return forecast methodologies. However, transaction costs  $v_t(\omega_{t+1})$  are not taken into account. Pettenuzzo and Ravazzolo (2016) compute portfolios (based on one risky asset) in a similar fashion.

Our methodology is flexible in the sense that it nests a variety of well-known approaches. Focusing on the expected utility after adjusting for transaction costs allows to adopt many of these methods to higher dimensions, as empirical considerations often point out, that neglecting transaction costs is a major concern (see DeMiguel et al. (2009)).

### 3. COMBINING PORTFOLIO STRATEGIES

Our optimization framework provides the tool to disentangle weight smoothness and predictive accuracy: at first, we compute return predictions under parameter and model uncertainty. Reducing turnover by weight smoothing is done in the subsequent step and should not be of any concern at this stage.

We propose a range of different predictive models to adjust for parameter uncertainty. Structural changes in the return distribution, time-varying parameters and the unknown structure of the DGP make it impossible to identify the main factors of the future return structure precisely. Instead, we combine the distinct models to improve our forecast abilities and claim that focusing

on a single strategy means to discard valuable information.<sup>9</sup> The investor chooses  $K$  distinct predictive return distribution models  $\Xi := \{\mathcal{M}_1, \dots, \mathcal{M}_K\}$ , either through personal preferences, data availability or theoretical considerations. In the advent of large amounts of data available the possibility to incorporate different sources of input is a valuable option. Over time, sequential learning about the parameters of the individual models takes place and in addition the investor decides how much her investment decision is driven by each of the individual models, conditional on the past performance. The reliability of a model may change over time, which is reflected in the sensibility of the portfolio weights with respect to updated beliefs in the individual models: predictive models which performed superior in the recent past obtain higher weights in the combination scheme.

Well-known approaches to combine different models are, among many others, Bayesian model averaging (see Hoeting et al. (1999)), optimal prediction pooling (Geweke and Amisano, 2011) and decision-based model combinations (Billio et al., 2013).

In general, when combining multivariate predictive densities we are interested in a time-series of combination weights  $c_t \in \mathbb{R}^K$ . These weights can be interpreted as discrete probabilities over the set of models  $\Xi$ . To justify the probabilistic interpretation of the combination scheme  $c_t$  it is required that all weights take positive values and add up to one:

$$c_t \in \Delta_{[0,1]^K} := \left\{ c \in \mathbb{R}^K : c_i \geq 0 \forall i = 1, \dots, K \text{ and } \sum_{i=1}^K c_i = 1 \right\}. \quad (10)$$

Conditional on the combination weights  $c_t$ , the predictive density of the combined models takes into account the predictions of each individual model  $\mathcal{M}_i$  to forecast returns of the next period: define the associated predictive return distributions  $\tilde{r}_{\mathcal{M}_i, t+1} \sim p_t(r_{t+1} | \mathcal{M}_i, \mathcal{F}_t, R_t)$ . To simplify notation we define the stacked  $NK$  vector of predictive distributions as

$$\tilde{r}_{t+1} := \text{vec}(\{\tilde{r}_{\mathcal{M}_1, t+1}, \dots, \tilde{r}_{\mathcal{M}_K, t+1}\}) \quad (11)$$

and denote with  $p_t(\tilde{r}_{t+1})$  its joint predictive density. Then, the predictive return distribution can

---

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

be rewritten as<sup>10</sup>

$$p_t(r_{t+1}|R_t, \mathcal{F}_t, \Xi) = \int \int p_t(r_{t+1}|c_{t+1}, \tilde{r}_{t+1}) p_t(c_{t+1}|\tilde{r}_{t+1}) p_t(\tilde{r}_{t+1}) d\tilde{r}_{t+1} dc_{t+1}. \quad (12)$$

This general notation nests a range of widely applied approaches: equation 12 allows for a stochastic interpretation of  $c_t$ , reflecting uncertainty with respect to the true value of the combination weights, but one can also regard the term  $p_t(c_{t+1}|\tilde{r}_{t+1})$  as deterministic. In this case, it remains to compute the value of  $c_t$  as a function of the predictive returns of the distinct models  $\tilde{r}_{t+1}$ . There are multiple ways to compute  $c_t$  as functions of the time-series of predictive returns of the model:

$$c_t = \underset{c \in \Delta_{[0,1]^K}}{\operatorname{argmax}} f(c, \tilde{r}_1, \dots, \tilde{r}_t). \quad (13)$$

Geweke and Amisano (2011) perform model combination by focusing on the past log predictive likelihood of the set of models, which is defined as

$$LS(R_t, \mathcal{F}_t, \mathcal{M}_k) = \sum_{i=1}^t \log p(r_i|R_{i-1}, \mathcal{F}_{i-1}, \mathcal{M}_k). \quad (14)$$

$p(r_t|R_{t-1}, \mathcal{F}_{t-1}, \mathcal{M}_k)$  is computed by evaluating

$$p(r_t|R_{t-1}, \mathcal{F}_{t-1}, \mathcal{M}_k) = \int p(r_t|R_{t-1}, \mathcal{F}_{t-1}, \Theta, \mathcal{M}_k) d\Theta \quad (15)$$

$$\approx \frac{1}{M} \sum_{m=1}^M p(r_t|\Theta^{(m)} \mathcal{M}, R_{t-1}, \mathcal{F}_{t-1}), \quad (16)$$

where  $M$  is the number of MCMC-draws. If the multivariate predictive density concentrates around the observed values, the log predictive likelihood is higher. Therefore, the log predictive likelihood of model  $\mathcal{M}_k$  reflects a measure of the predictive accuracy. The time-varying combination weights  $c_t$  can be chosen to put weights on the models in a way to maximize the log-predictive likelihood of the combined model:

$$c_t^* = \underset{c \in \Delta_{[0,1]^K}}{\operatorname{argmax}} \sum_{i=t-h}^t \log \left[ \sum_{k=1}^K c_k p(r_i|R_{i-1}, \mathcal{F}_{i-1}, \mathcal{M}_k) \right], \quad (17)$$

where  $h$  defines the horizon length. Durham and Geweke (2014) provide empirical evidence for the benefits in terms of predictive accuracy for asset price predictions based on prediction pooling.

---

<sup>10</sup>The last term of this expression,  $p_t(\tilde{r}_{t+1})$ , can be evaluated by the MCMC sequences of the predictive return distribution of each model  $\mathcal{M}_1, \dots, \mathcal{M}_K$ . Regions which are more likely within the joint predictive return distribution affect the predictions more than regions which are viewed unlikely to occur.

Choosing  $c_t$  as a function of the predictive return distributions of the set of models  $\Xi$  also allows to focus on combination schemes based on the past portfolio performance: at every point in time the models  $\mathcal{M} \in \Xi$  generate portfolio weights  $\{\omega_t^1, \dots, \omega_t^k\} \in \mathbb{R}^{N \times k}$  as solutions to the maximization problem (EU). Therefore,  $c_t$  can also be computed choosing, for instance,

$$c_t = \arg \max_{c \in \Delta_{[0,1]^K}} (1 - \lambda) \sum_{i=t-h}^t \lambda^{t-i} c U_\gamma(\omega'_{i,k} r_{i+1} - v(\omega_i, \omega_{i-1,+})). \quad (18)$$

In this case  $c_t$  puts more importance on the recent past and ignores events which occurred more than  $h$  periods ago. Given model  $\mathcal{M}_k$  performs good and generated high utility in the past,  $c_t$  is high.  $c_t$  increases if the performance of a portfolio allocation strategy increases. Billio et al. (2013) embed such a decision-based approach in a Bayesian setting by assuming a Markovian process for  $c_t$ , adjusting for the case of misspecification of the weighting scheme.

Whereas optimal prediction pooling is driven solely by the predictive performance of the models, the crucial measure for investors to evaluate the allocation strategies is the predictive out-of-sample portfolio return

$$\tilde{r}_{t+1}^{\text{pf}} := \hat{\omega}'_{t+1} \tilde{r}_{t+1} - v_t(\hat{\omega}_{t+1}, \hat{\omega}_{t,+}). \quad (19)$$

In general, there can occur a discrepancy between predictive accuracy and portfolio performance due to the distortion through transaction costs. Even if a model generates very precise predictions regarding the returns of the next time-period, it might still be beneficial in terms of realized portfolio returns to smooth rebalancing to avoid high fees or to generate price impact by the investors own trade. Our optimization framework accounts for transaction costs and therefore we do expect that combinations based on past portfolio performance do not differ in a significant manner from combinations which focus on the past predictive performance.

## 4. HIGH-FREQUENCY AND LOW-FREQUENCY BASED PREDICTIVE RETURN DISTRIBUTIONS

### 4.1. A VAST-DIMENSIONAL WISHART MODEL FOR BLOCKED REALIZED KERNELS

Intra-daily messages are far more informative about the current level of volatility than a single observation of the squared daily return. Realized measures of volatility based on HF data have been shown to provide accurate estimates of daily covariances.<sup>11</sup> We expect HF data to provide an economic significant benefit when used as input to the portfolio allocation procedure, as it

---

<sup>11</sup> See e.g. Andersen and Bollerslev (1998), Andersen et al. (2003) and Barndorff-Nielsen et al. (2009).

captures current market movements much faster than estimators based on a long time-series of past daily observations. This belief is in line with the findings of Hautsch et al. (2015) and Liu (2009).

Estimating covariances based on intra-daily messages suffers from microstructure noise which requires filtering methodologies to retain estimates of the variable of interest. In addition, trading arises in irregular time-patterns. The multivariate realized kernel estimator of Barndorff-Nielsen et al. (2011) is the first to address market microstructure and asynchronous price observations while guaranteeing consistency and positive-semi-definiteness. In order to obtain forecasts of covariances based on high-frequency data we employ block realized kernel (BRK) estimates as proposed by Hautsch et al. (2012) to estimate the quadratic variation of the price process based on discretely sampled and noisy price observations.

Synchronizing is obtained via refresh time sampling which discards many observations if the asset horizon is large. Hautsch et al. (2012) propose a methodology to reduce dropout by composing blocks of asset clusters with similar liquidity structure and computing realized kernel estimators block-wise. This increases the precision of the estimates as less observations are discarded.

We group assets conditional on the number of daily messages. Hereby, the periods between refresh times for the assets belonging to the group with highest liquidity are generally smaller. We denote the set of indices of the assets associated with block  $b \in 1, \dots, 6$  by  $\mathcal{J}_b$ . For each asset  $i$ ,  $\tau_{t,l}^{(i)}$  denotes the time-stamp of message  $l$  on day  $t$ . Refresh times are defined as

$$r\tau_{t,1}^b := \max_{i \in \mathcal{J}_b} \left\{ \tau_{t,1}^{(i)} \right\}, \quad r\tau_{t,l+1}^b := \max_{i \in \mathcal{J}_b} \left\{ \tau_{t,N^{(i)}(r\tau_{t,l}^b)}^{(i)} \right\} \quad (20)$$

where  $N^{(i)}(\tau)$  denotes the number of price observations of asset  $i$  before time  $\tau$ . The definition of refresh times synchronizes the observations:  $r\tau_{t,i}^b$  denotes the point in time where all of the assets belonging to group  $b$  have been traded at least once since the last refresh time  $r\tau_{t-1,i}^b$ . Based on  $r\tau_{t,l+1}^b$ , synchronized returns are obtained as  $r_{t,l}^{(i)} := p_{r\tau_{t,l}^b}^{(i)} - p_{r\tau_{t,l-1}^b}^{(i)}$  with  $p_{r\tau_{t,l}^b}^{(i)}$  the log mid-quote of asset  $i$  at time  $r\tau_{t,l}^b$ . For each block, the multivariate realized kernel (Barndorff-Nielsen et al., 2011) is defined as

$$K_t^b := \sum_{h=-H_t^b}^{H_t^b} k\left(\frac{h}{H_t^b+1}\right) \Gamma_t^{h,b}, \quad (21)$$

where  $k(\cdot)$  is the Parzen kernel and  $\Gamma_t^{h,b}$  is the  $h$ -lag auto-covariance matrix of the assets belonging to block  $\mathcal{J}_b$ . The estimates of the correlation of assets in block  $b$  takes the form

$$\hat{H}_t^b = \left(V_t^b\right)^{-1} K_t^b \left(V_t^b\right)^{-1} \quad V_t^b = \text{diag} \left[ K_t^b \right]^{1/2}. \quad (22)$$

To obtain the  $N \times N$  correlation matrix  $\hat{H}_t$ , the blocks  $\hat{H}_t^b$  are stacked as described in Hautsch et al. (2012). The variances  $\sigma_{t,i}^2$  of each asset are computed based on the complete set of daily observations by computing standard realized kernel estimates. Finally, the variance-covariance matrix based on HF data is computed by

$$\hat{\Sigma}_t^{\text{BRK}} = \text{diag}(\sigma_{t,1}^2, \dots, \sigma_{t,N}^2)^{1/2} \hat{H}_t^b \text{diag}(\sigma_{t,1}^2, \dots, \sigma_{t,N}^2)^{1/2}. \quad (23)$$

Although implying efficiency gains, block-wise estimation comes at the cost of positive semi-definite and well-conditioned variance-covariance estimates. Eigenvalues which are close to zero or even negative have to be adjusted in order to retain well-defined estimators. We employ cleaning following Laloux et al. (1999). This approach compares the eigenvalues of the estimates with the derived distribution of eigenvalue given the processes would be independent and is derived from random matrix theory. If the empirical eigenvalues are close to zero they are identified as not carrying information on cross-sectional dependencies. Inflating these values recovers well-conditioned matrices without significantly losing information.

The estimates  $\hat{\Sigma}_t^{\text{BRK}}$  represent the realized volatilities and covariances within trading day  $t$ . To use these estimates for forecasting purposes it is required to parametrize a suitable return distribution which is driven by the dynamics of  $\hat{\Sigma}_t^{\text{BRK}}$ . In a first step we ignore parameter uncertainty and proceed with a random-walk forecasting procedure by assuming  $p_t(r_{t+1} | \hat{\Sigma}_t^{\text{BRK}}) \sim N(0, \hat{\Sigma}_t^{\text{BRK}})$ . This parametrization implies that the current state of market, represented by  $\hat{\Sigma}_t^{\text{BRK}}$  also represents the market volatility of tomorrow. However, although this approach makes use of highly informative data, ignoring parameter uncertainty with respect to the covariances of tomorrow affects the predictive accuracy.

We extend this model by embedding our approach in a Bayesian setting in the spirit of Jin and Maheu (2013). To capture uncertainty with respect to the parameters of the DGP, we model integrated volatility as a Wishart distribution, which represents a multivariate extension of the normal-inverse-gamma approach, applied for example by Andersson (2001), Jensen and Lunde (2001), Forsberg and Bollerslev (2002) and Barndorff-Nielsen (1997).

The Wishart distribution has the advantage that its support consists of the set of nonnegative-definite matrices. Regularization of the posterior estimates to obtain interpretable results is therefore not necessary anymore. We model the dynamics of the predicted return process conditional on the latent covariance  $\Sigma_t$  as multivariate Gaussian

$$\mathcal{L}(r_{t+1} | \Sigma_{t+1}) \sim N(0, \Sigma_{t+1}). \quad (24)$$



To justify our likelihood assumption, we rely on the mixture-of-distribution hypothesis which starts from the premise that the distribution of discretely sampled returns (such as daily ones) conditional on latent information is Gaussian (Clark, 1973). From a Bayesian perspective we consider the latent variable  $\Sigma_{t+1}$  as uncertain and assume

$$\Sigma_{t+1}|\kappa, S_t \sim IW(\kappa, S_t) \quad (25)$$

where  $\kappa > N$  and  $S_t = f(\hat{\Sigma}_1^{BRK}, \dots, \hat{\Sigma}_t^{BRK})$  is measurable with respect to  $\mathcal{F}_t$ . We employ

$$S_t = \hat{\Sigma}_t^{BRK}. \quad (26)$$

Combining the distributional assumptions yields the implied posterior distribution:

$$\pi(\Sigma_{t+1}|R_t, \mathcal{F}_t) \sim IW\left(\sum_{i=1}^t (R_i - \bar{R})'(R_i - \bar{R}) + \kappa S_t, t + \kappa\right). \quad (27)$$

This distribution exhibits the posterior mean

$$E(\Sigma_{t+1}|R_t, \mathcal{F}_t) = \frac{1}{t + \kappa - N - 1} \sum_{i=1}^t (R_i - \bar{R})'(R_i - \bar{R}) + \kappa S_t \quad (28)$$

$$= \frac{t - N - 1}{t + \kappa - N - 1} \hat{\Sigma}_t + \frac{\kappa}{t + \kappa - N - 1} S_t. \quad (29)$$

This specification is sensible to parameter uncertainty regarding the latent factor  $\Sigma_{t+1}$  and adjusts depending on the choice of the hyperparameter  $\kappa$ .

#### 4.2. MULTIVARIATE FACTOR STOCHASTIC VOLATILITY MODELS

Return time series often exhibit time-varying volatilities and co-movements among the distinct assets. Often referred to as the curse of dimensionality, modeling this behavior with many assets requires a trade-off between flexibility of feasibility of computation. Multivariate stochastic volatility models are one way to impose a structure which is flexible enough to capture these stylized facts and to keep the number of parameters on a reasonable low level. Han (2006) derives portfolio weights incorporating time-varying forecasts of the covariance matrix in the spirit of stochastic volatility models and obtains strong portfolio performances. This is in line with the recent findings of Johannes et al. (2014), stating evidence for return predictability provided time-varying volatility and estimation risk is adjusted for.

Lopes et al. (2016) propose an efficient way to compute Bayesian estimates of dynamic covariance matrices in multivariate time series through a stochastic volatility factor model in the spirit

of Shephard (1996), Jacquier et al. (2002) and Chib et al. (2006). The factor model facilitates the numerical burden and ensures flexibility of the coefficients to capture volatility clustering and volatility co-movement. Time-varying variances are allowed for both the  $N$  idiosyncratic innovations as well as  $j$  factors, depending on  $N + j$  latent volatilities  $h_t = (h_{1,t}, \dots, h_{N+j,t})$ . Our implementation follows the structure outlined in Lopes et al. (2016) and takes the form

$$r_t = \Lambda V(h_t)^{1/2} \zeta_t + Q(h_t)^{1/2} \varepsilon_t. \quad (30)$$

$\Lambda \in \mathbb{R}^{N \times j}$  is the unknown factor loadings matrix,  $Q(h_t) = \text{diag}(\exp(h_{1,t}), \dots, \exp(h_{N,t})) \in \mathbb{R}^{N \times N}$  and  $V(h_t) = \text{diag}(\exp(h_{N+1,t}), \dots, \exp(h_{N+j,t})) \in \mathbb{R}^{j \times j}$  contain the factor variances. The variances  $h_t$  are modeled as latent variables following AR(1)-processes

$$h_{i,t} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sigma_i \eta_{i,t} \quad i = 1, \dots, N + j. \quad (31)$$

Hereby, the initial state  $h_{i0}$  is unknown. The innovations  $\varepsilon_t, \zeta_t, \eta_t$  are assumed to follow independent standard normal distributions. This model implies

$$\text{cov}(r_t | h_t) = \Sigma_t(h_t) = \Lambda V_t(h_t) \Lambda' + Q_t(h_t). \quad (32)$$

Appealingly, instead of computing  $(N + 1)N/2$  correlation parameters, co-movements between assets are affected only through the factor loadings  $\Lambda$  due to the parsimonious structure of the diagonal matrices  $Q$  and  $V$ . The priors for the univariate stochastic volatility processes are chosen independently

$$p(\mu_i, \phi_i, \sigma_i) = p(\mu_i) p(\phi_i) p(\sigma_i) \quad (33)$$

The level  $\mu_i$  is equipped with a normal prior, the persistence parameter  $\phi_i$  is chosen such that  $(\phi_i + 1)/2 \sim B(a_0, b_0)$ , and  $\sigma_i^2 \sim G\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$ . For each element of the factor loadings matrix, a zero-mean Gaussian distribution is chosen.

The SV model has been shown to provide good results when it comes to modeling covariances of financial time series. Therefore it serves as a natural candidate in the portfolio allocation context.<sup>12</sup>

#### 4.3. LEDOIT-WOLF COVARIANCE REGULARIZATION

One standard approach to predict return volatilities is the sample covariance estimator

$$S_t := \frac{1}{t-1} \sum_{i=t-h}^t (r_i - \hat{\mu}_t)(r_i - \hat{\mu}_t)' \quad (34)$$

<sup>12</sup>The implementation is based on the R package **factorstochvol** (Kastner, 2016). The algorithm to sample from the posterior distribution is explained in detail in Lopes et al. (2016).

with  $\hat{\mu}_t := \frac{1}{h-1} \sum_{i=t-h}^t r_i$  and estimation window size  $h$ . However, often it is preferred to deviate from the consistent estimator  $S_t$  and instead to choose a biased but less volatile estimator. The shrinkage variant of the covariance-estimator of Ledoit and Wolf (2004) takes the form

$$\hat{\Sigma}^{t, \text{Shrink}} = \hat{\delta} F + (1 - \hat{\delta}) S_t \quad (35)$$

where  $\hat{\delta}$  minimizes the Frobenius norm between the sample constant correlation matrix  $F$  and the sample covariance estimator  $S_t$ .  $F$  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  $j$  column of the sample covariance matrix  $S_t$ . The average sample correlations are given by  $\bar{\rho} := \frac{2}{(N-1)N} \sum_{i=1}^N \sum_{j=i+1}^{N-1} \hat{\rho}_{ij}$ .  $F$  is defined as  $F_{ij} = \bar{\rho} \sqrt{\hat{\rho}_{ii}\hat{\rho}_{jj}}$ .<sup>13</sup> Modeling the time-series of future returns is performed by assuming  $p_t(r_{t+1} | \hat{\Sigma}^{t, \text{Shrink}}) \sim N(0, \hat{\Sigma}^{t, \text{Shrink}})$ .

## 5. DATA

We consider an investment strategy with daily readjustment for  $N = 100$  stocks traded at NASDAQ during 27th of June 2007 until 27th of June 2016. After excluding weekends and holidays, the time series of returns covers 2265 trading days. The 100 stocks are sampled randomly from the total population of available tickers. Companies whose stocks were not traded during the complete time period are excluded from the analysis. All the computations are performed after adjusting for stock splits and dividends. Our dataset of further explanatory variables  $\mathcal{F}_t$  contains intra-daily message data, extracted from LOBSTER.<sup>14</sup> For each asset, LOBSTER reconstructs the full limit order book within each trading day. We concentrate on forecasting volatilities based on realized kernels and only investigate the time-series of mid-prices.

Panel a) of Figure 1 visualizes the cross-sectional average of realized volatilities for each trading day. Volatile periods, such as the financial crisis, are clearly represented in the sample. Furthermore, persistence and time-varying properties of assets volatilities are present. Panel b) of Figure 1 shows the average of absolute correlation computed using block realized kernel estimates on a daily basis. The figure underlines the time-varying properties of correlations among financial assets. Furthermore, persistence in co-movement among the selected assets can be detected. Figure 2 visualizes the empirical correlation matrices based on daily observations at three different points in time. Panel a) belongs to the estimation horizon reaching from June 2007 until June 2009 and exhibits the strongest correlation among the assets compared to calmer market

<sup>13</sup>The formula for the shrinkage intensity  $\hat{\delta}$  can be found in Appendix B in Ledoit and Wolf (2004).

<sup>14</sup>See <https://lobsterdata.com/>.

periods such as the time between June 2013 and June 2016. Blue colors indicate positive correlations. The brighter the color, the less pronounced the correlation. During the complete time horizon the assets always exhibit positive correlation amongst each other. This is probably due to the common market place they share. In terms of portfolio allocation the positive correlation patterns should make it harder to come up with strong diversification effects due to the absent of natural hedging instruments. We aim at constructing predictive models capturing these essential properties of time-series. Subsequently, we expect portfolio performance gains by exploiting the informational content inherent in the predictions.

## 6. EMPIRICAL FINDINGS

### 6.1. EVALUATION OF PREDICTIVE PERFORMANCE

We disentangle between the (i) predictive accuracy and (ii) portfolio performance of the models taken into consideration. First, we visualize the substantial differences of the models used in our setting for a particular example and provide some insights on the predictions made by the distinct models: based on the data available up to point in time  $t$ , we sequentially compute the predictive distribution of the naive portfolio returns ( $\omega_{\text{naive}} := \frac{1}{N} \iota$ ) at day  $t+1$ . This is done using MVSFV, the BRK approach based on HF-data and the Gaussian setting with Ledoit-Wolf regularization,  $\hat{\Sigma}_t^{\text{LW}}$ . The HF-data is exploited by first computing the BRK approach ignoring parameter uncertainty as we assume multivariate Gaussian distributed returns  $r_{t+1} \sim N(0, \hat{\Sigma}_t^{\text{BRK}})$ . In addition, we forecast returns by sampling from the predictive distribution using the covariance-matrix resulting from the Bayesian Normal-Wishart approach (see Equation (27)).

For each of these models we generate 50.000 sample draws according to Equation (8). First, we generate samples from the posterior distribution of the parameters, then we generate  $N = 100$  dimensional predictions  $r_{t+1}$  and in a subsequent step we project these vectors into one dimension by computing  $\omega'_{\text{naive}} r_{t+1}$ . This procedure generates samples from the predicted return distribution of the naive portfolio.

Figure 3 visualizes the resulting 95% credible regions: the purple dots indicate the true return of the naive portfolio at day  $t+1$  (the value our models aim to predict), whereas the blue lines show the predictive interval computed with the stochastic volatility factor model. The smooth green lines correspond to the Ledoit-Wolf regularized variance-covariance estimation based on the past 250 trading days. The yellow line captures the predictions based on the HF-data computed with BRK. The red line corresponds to our Wishart-model for blocked realized kernels. The figure

indicates the superiority of the stochastic volatility factor model compared to the (static) Ledoit-Wolf approach. MVSFV and BRK both seem to capture changes in the market environment very fast.<sup>15</sup>

The results of a more in-depth analysis of the predictive performance are presented in Table 2 to underline the predictive performance of the applied approaches in a more comprehensive manner: we compute the log posterior predictive likelihood for MVSFV, Ledoit-Wolf regularization and the BRK approaches as shown in Equation (15). The log posterior predictive likelihood indicates how much probability mass the individual predictive distribution assigned to the observed outcomes. We compute summary statistics of the time-series of likelihoods to evaluate the predictive performance: the stochastic volatility factor model outperforms its peers in terms of the mean posterior log-likelihood obtained in the sample.

Forecasting large-dimensional time-series based on a rolling-window estimate of the sample variance-covariance matrix is dominated by imposing additional regularization in the form of Ledoit-Wolf shrinkage. However, these models perform rather poor compared to the dynamic implementations based on stochastic volatility and the HF-data based estimates.

The forecasts based on the BRK-estimates neglecting parameter uncertainty do not exhibit fat tails and cannot be used as reliable predictive models in our setting. However, incorporating parameter uncertainty by marginalizing out the uncertain parameters improves predictive accuracy compared to the BRK estimates. Adjusting for parameter uncertainty by using the dynamic specification proposed in Equation (27), the forecast quality of the HF-data based estimates increases considerably: implied fat tails decrease the variance of the predictive log-likelihood and increase the mean precision, denoted in the second-to-last column of Table 2.

The last row of Table 2 denotes the fraction of days through the sample, on which the individual models generates the best prediction in terms of predictive log-likelihood. Apparently, the Normal-Wishart parametrization outperformed its peers on 41% of the days. This provides further evidence on the importance of HF-data to predict market movements.

We also adjust for model uncertainty and compute predictions by mixing the individual models: computing a time-series of combination weights  $c_t$  in the spirit of Geweke and Amisano (2011) allows to put weight on each of the models according to the past predictive accuracy in terms of the predictive log-likelihood. The sequentially updated combination weight vector  $c_t$  is plotted in Figure 4. Each time-series of Panel a) corresponds to one of the 5 implemented mod-

---

<sup>15</sup> The extreme spike corresponding to the BRK estimates on May 2010 corresponds to the severe intra-daily volatility caused by the Flash Crash which did not only affect the S&P 500 but also affected many stocks, leading to extreme estimates for this day. A similar extreme event can be observed in August 2015.

	MVSFV	Sample	LW	BRK	BRK <sub>NW</sub>	Mixing
Mean	309.22	276.57	300.90	224.71	303.78	<b>323.03</b>
Standard deviation	24.42	70.82	37.75	98.12	32.02	<b>20.58</b>
Range	185.38	615.73	314.6	915.34	211.92	152.73
% best prediction	0.27	0.16	0.16	0.01	<b>0.41</b>	-

Table 2: Predictive accuracy of the predictive models. At every period in time  $t$  the past returns  $R_t$  are used to compute the predictive distribution of the models. *Mean* denotes the average posterior predictive log-likelihood based on 1904 computed forecasts for the returns of the 100 assets. *Standard deviation* and *Range* summarize the dispersion of these estimates. The last row denotes the fraction of days where the individual models performed best among their peers in terms of predictive log-likelihood. *MVSFV* denotes a factor model with  $j = 2$  factors. *Sample* covariance is computed by implying constant covariance matrix  $r_{t+1} \sim N(0, S)$  with  $S$  being the sample variance-covariance matrix based on an estimation window of size  $h = 250$  days, Ledoit-Wolf regularization is based on the regularized variance-covariance forecast  $r_{t+1} \sim N(0, \hat{\Sigma}_{\text{Shrink}})$ . *BRK* corresponds to covariance estimates based on HF-data, obtained by computed block-realized kernel estimates with 6 blocks. *BRK<sub>NW</sub>* corresponds to the inverse Wishart specification proposed in Section 4.1. *Mixing* corresponds to predictions created by combining the individual models in the spirit of Geweke and Amisano (2011).

els. The BRK approach without adjusting for parameter uncertainty is dominated by the Normal-Wishart approach and does not receive any weight. However, the dynamic of the remaining approaches is time-varying and exhibits interesting patterns: throughout the whole sample period, most of the weight is put on the HF-data, followed by the stochastic factor volatility model. The weights of the Ledoit-Wolf approach as well as the sample variance-covariance counterpart never exceed 20%. Panel b) of Figure 4 denotes the model weights conditional on the data source: the red line corresponds to the total fraction assigned to models based on *HF* data, whereas the blue line shows the cumulative sum of model probabilities of all models based on daily data as input source. Apparently, the relative importance of *HF* data is changing over time.

The last column of Table 2 reveals the impressive impact of model combination for predictions: forecasting high-dimensional return time-series based on iteratively updating combination scheme  $c_t$  and mixing the models accordingly results in predictive log-likelihoods which are

superior on average and exhibit the lowest standard deviation among all individual models.

However, superior predictive accuracy is not equivalent to superior portfolio performance. The volatility of the predictive credible regions visualized in Figure 3 underlines that an investor not adjusting for transaction costs *ex ante* may rebalance her portfolio unnecessarily often if she relies on the predictions based on the stochastic volatility factor model. Although the underlying predictions may be quite accurate, transaction costs can easily offset this advantage compared to using the rather smooth prediction of the Ledoit-Wolf regularized forecasts.

## 6.2. EVALUATION OF PORTFOLIO PERFORMANCE

Besides the statistical performance of our prediction models, we aim at investigating the benefits of our approach in terms of economical significance. In order to do this we perform a portfolio allocation horse-race to compare the performance of our approaches. First, we fix the asset horizon and the set of  $k$  models  $\Xi$ . Then, at each time period, the investor uses the available information to update her beliefs about the parameters and state variables. Subsequently, based on the estimates, she generates predictions for the returns of tomorrow and allocates her wealth accordingly by solving Equation (EU). After holding the assets for a day, she realizes the gains and losses, updates posterior distribution and recomputes optimal portfolio weights. This procedure is repeated for each period and allows to analyze the time-series of obtained out-of-sample returns  $r_{t+1}^k = \omega'_{k,t+1} r_{t+1}$  as well as the out-of-sample returns net of transaction costs for each of the implemented models  $k$ .

We focus our interest to the out-of-sample portfolio returns after taking into account transaction costs of the form

$$v_t(\omega_{t+1}, \omega_t) = \frac{50}{10.000} \left\| \omega_{t+1} - \frac{w'_t r_t}{\|w'_t r_t\|_1} \right\|_1 \quad (36)$$

as suggested by DeMiguel et al. (2009). Hereby, the term  $\frac{w'_t r_t}{\|w'_t r_t\|_1}$  denotes the portfolio weights at point in time  $t$  *immediately before* readjustment. Therefore, the portfolio returns of strategy  $k$  are defined as

$$r_{t+1}^{k,\text{OOS}} := r_{t+1}^k - v_t(\omega_{t+1}^k). \quad (37)$$

Based on the out-of-sample returns after transaction costs we evaluate the mean and the variance

$$\mu(r_T^{k,\text{OOS}}) := \frac{1}{T} \sum_{t=1}^T r_{t+1}^{k,\text{OOS}}, \quad \sigma(r_T^{k,\text{OOS}})^2 := \frac{1}{T-1} \sum_{t=1}^T \left( r_{t+1}^{k,\text{OOS}} - \mu(r_T^{k,\text{OOS}}) \right)^2. \quad (38)$$

Furthermore, the Sharpe Ratio is evaluated as

$$SR_T^k := \frac{\mu(r_T^{k, \text{OOS}})}{\sqrt{\sigma(r_T^{k, \text{OOS}})^2}}. \quad (39)$$

In addition we examine several characteristics related to the portfolio turnover and the composition of the portfolio allocation. The average weight concentration is given by

$$pc_k := \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \omega_{t,n}^2. \quad (40)$$

The average size of the short positions is evaluated by computing

$$sp_k = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N |\omega_{n,t}| \mathbb{1}\{\omega_{n,t} < 0\}. \quad (41)$$

The procedure of our empirical simulation is summarized in Table 3, taking into account the combination strategies presented. After choosing  $N$  assets and a set of models  $\Xi$ , at each point in time the investor observes the available data  $(R_t, \mathcal{F}_t)$ . Based on the observation the optimization framework stated in Equation EU is solved for each model  $\mathcal{M}_k$  after updating the parameters and computing the predictive return distribution  $p_t(r_{t+1}|\mathcal{M}_k)$ . The time-series of past realized returns  $\{r_i^{k, \text{OOS}}\}_{i=1, \dots, t-1}$  is used to compute the combination weights  $c_t$ . In the last step, the predictive return distribution after accounting for model uncertainty,  $p_t(r_{t+1}|\Xi)$  is used as input to solve Equation EU again. The resulting allocation vector  $\omega_{t+1}^*$  determines rebalancing.

- |            |  |
|------------|--|
| At $t = 0$ | Define investment horizon $N$<br>Define set of $K$ models $\Xi$  |
| At $t > 0$ | Investor observes past returns $R_t$ and $\mathcal{F}_t$<br>For each model $\mathcal{M}_k \in \Xi$ compute the posterior predictive $p_t(r_{t+1} \mathcal{M}_k)$<br>Solve (EU) for $\omega_{\mathcal{M}_k, t+1}$<br>Based on $R_t$ and $p_t(r_{t+1} \mathcal{M}_k)$ compute value of objective function $\xi_{t,k}$<br>$c_t$ serves as input to compute $p_t(r_{t+1} \Xi)$ .<br>Investment decision $\omega_{t+1}^*$ is obtained by solving (EU) based on $p_t(r_{t+1} \Xi)$ . |

Table 3: Initial decisions and procedure to compute portfolio weights taking into account model uncertainty, parameter uncertainty and transaction costs within our framework.

We implement portfolio allocation strategies based on the predictive models described above and the proposed mixing strategies. The forecasting and optimization procedure requires large



computing resources. The computational result have been achieved in part using the Vienna Scientific Cluster.<sup>16</sup> Additionally, we implement the naive portfolio allocation. Investing the wealth into equal parts  $\omega_{\frac{1}{N}} := \frac{1}{N} \mathbf{1}$  is not prone to any estimation error. Furthermore, the transaction costs associated with this strategy are considerably low due to the persistence in the weights. Among others, DeMiguel et al. (2009) shows that the naive portfolio often outperforms more sophisticated strategies in terms of out-of-sample returns.

The Bayes-Stein estimator of Jorion (1986) tackles model and parameter uncertainty by modifying shrinkage parameters to minimize the effect on utility loss of errors. This method serves as a competitor and is also implemented. Details with respect to the implementation are given in the Appendix. Furthermore, the global minimum variance portfolio is computed based on the Ledoit-Wolf Regularization approach which is the solution to the optimization problem

$$\omega_{\text{GMVP}} := \underset{\omega \in \mathbb{R}^N: \mathbf{1}'\omega=1}{\operatorname{argmin}} \omega' \Sigma^{LW} \omega.$$

Additionally we compare the results to the GMVP portfolio with a no-short sale constraint as proposed by Jagannathan and Ma (2003). A summary of the distinct strategies implemented is listed in Table 6. Table 4 summarizes the portfolio performances of the distinct models implemented. The evaluation of portfolio returns is computed both *before* and *after* adjusting for transaction costs. The main results can be summarized as follows:

1. Without regularizing turnover, portfolio allocation in high dimensions is doomed to fail.
2. Our approach transforms good predictions into strong portfolios.
3. Model combination increases the portfolio performance considerably.

#### 6.2.1. BENEFITS OF REGULARIZATION OF TRANSACTION COSTS

First, we compare the allocation of an investor ignoring transaction costs *ex ante* (by setting  $\nu(\omega) = 0$  when computing optimal weights  $\omega^*$ ) with the optimization result if we embed regularization according to equation EU. The associated allocation strategies are depicted as  $\omega_{\text{LW,unrest}}$  and  $\omega_{\text{MVSFV,unrest}}$ .<sup>17</sup> The results clearly show that penalizing turnover is crucial to obtain satisfying portfolio performances: optimal weights suffer from high turnover and the resulting portfolio returns are consumed by the transaction costs. After adjusting for transaction costs, neither

<sup>16</sup>All computations are done with **R**. The source code is available upon request.

<sup>17</sup>Equivalent computations have been performed for the remaining models as well but are not displayed. The resulting portfolio performance look similar, none of the unrestricted allocation strategies performs well *after adjusting for transaction costs*.

	Before transaction costs			After transaction costs			Neg. W.	$pc$
	Mean	SD	SR	Mean	SD	SR		
$\omega_{LW, \text{unrest}}$	10.3	<b>10.54</b>	0.977	-3.2	10.52	-	1.11	0.49
$\omega_{MVSFV, \text{unrest}}$	13.6	12.89	1.055	-89.73	13.21	-	0.94	0.44
$\omega_S$	14.2	13.37	1.062	14.1	13.45	1.055	0.13	0.24
$\omega_{LW}$	13.8	13.41	1.029	13.7	13.39	1.023	0.13	0.23
$\omega_{BRK}$	13.5	15.71	0.859	13.4	15.61	0.852	0.15	0.20
$\omega_{BRK, NW}$	14.4	15.02	0.959	14.2	15.09	0.941	0.08	0.13
$\omega_{MVSFV}$	<b>16.6</b>	14.02	1.184	<b>16.4</b>	14.04	1.168	0.41	0.54
$\omega_{\text{Mixing}}$	16.2	13.62	<b>1.189</b>	16.1	13.70	<b>1.174</b>	0.18	0.35
$\omega_{GMVP}$	8.9	11.08	0.802	- 11.4	<b>10.01</b>	-	1.50	0.59
$\omega_{GMVP \text{ w/o short}}$	9.9	11.07	0.894	7.0	11.09	0.631	0.00	0.45
$\omega_{BS}$	5.09	12.82	0.397	-22.26	13.84	-	1.21	0.90
$\omega_{\text{Naive}}$	12.0	17.59	0.682	10.9	17.66	0.617	0.00	0.01

Table 4: Results of the empirical horse-race after out-of-sample based on 1904 trading days.

Transaction costs are imposed as proportional to the  $L_1$  norm of rebalancing (denoted in Equation 36).  $SR$  denotes the (annualized) out-of-sample Sharpe ratio of the individual strategies.  $Neg. W.$  is the average of the absolute value of negative portfolio weights.  $pc$  is the portfolio concentration, computed as the average  $L_2$  norm of the portfolio weights. A description of the trading strategies is provided in Table 6.

$\omega_{LW, \text{unrest}}$  nor  $\omega_{MVSFV, \text{unrest}}$  create value to the investor, although the predictive performance of these models seemed rather promising.

However, adjusting for transaction costs changes the picture: the implemented strategies perform well and all of them outperform competitors in terms of out-of-sample the Sharpe Ratios. The models perform stronger than the naive portfolio, which is remarkable in the case of  $N = 100$  assets. Imposing 'wrong' restrictions such as the no-short sale constraint for the global minimum variance portfolio ( $\omega_{GMVP \text{ w/o short}}$ ) does not seem to be able to reach the strong performance of our models neither. To give some intuition on the effect of turnover penalization, we plot the daily turnover of the individual strategies in Figure 6b. The Figure reveals the Lasso-like behavior of the optimization strategy: whereas the no-short sale constrained GMVP allocation strategy

does not penalize small transactions and instead exhibits a continuous flow of orders and therefore also transaction costs, the penalized strategies *MVSFV* and *BRK* exhibit a completely different trading pattern.<sup>18</sup> By adjusting updated information and capturing transaction costs within the optimization procedure, the investor implicitly implements a buy-and-hold strategy whereas trading only occurs if the sequentially updated beliefs about the parameters of the DGP propose a substantial different allocation than represented by her current allocation.

In line with our findings with respect to the accuracy of the predictors, adjusting for parameter uncertainty of the estimations based on HF data in the spirit of Equation (27) increases the portfolio performance. The obtained Sharpe ratio after transaction costs of the Normal-Wishart approach  $\omega_{BRK,NM}$  exceeds the 'naive' approach using the BRK estimates within a simple Gaussian framework.

### 6.2.2. OUT-OF-SAMPLE PERFORMANCE OF MODEL MIXING PROCEDURES

The predictive performance of our mixing procedure reveals that adjusting for model uncertainty provides a benefit when it comes to forecasting. The results of the associated portfolio strategy  $\omega_{Mixing}$  underline that the same holds for allocation strategies. Mixing the forecasts of the individual models based on their past predictive accuracy and using the resulting predictive return distribution  $p_t(r_{t+1}|\Xi) = \sum_{k=1}^K p_t(r_{t+1}|\mathcal{M}_k)c_{t,k}$  as input to our optimization framework generates portfolio weights which increase the performance even further: the obtained Sharpe ratio (before and after transaction costs) is highest among all implemented methods for our data set. This provides further evidence on the benefits of adjusting for model uncertainty to account for structural breaks and time-varying dynamics in the DGP of future returns.

### 6.2.3. PORTFOLIO BOOTSTRAPPING

Obviously, a crucial input parameter of an portfolio allocation strategy is the set of investment possibilities  $n \in \{1, \dots, N\}$ . In order to underline the robustness of our results, we implement a portfolio bootstrap by repeating the horse-race explained above 100 times by repeatedly sampling  $N = 100$  assets and performing the optimization procedure again. The results are summarized in Table 5. Computing allocations based on the stochastic volatility factor model (by simply taking the posterior median of the variance-covariance matrix and neglecting higher moments) outperforms the competing models when adjusting for transaction costs. The results seem robust over the choice of the investment horizon and on average the Multivariate Stochastic Factor

---

<sup>18</sup>Note that the naive portfolio also exhibits turnover, as readjustment occurs at every period.

Volatility model generates the highest returns after adjusting for transaction costs. To underline this statement, the corresponding cross-section of Sharpe ratios (after transaction cost) is visualized in Figure 7. Each box plot corresponds to the realized out-of-sample Sharpe ratios for the distinct methods implemented. Evidently, our optimization procedures are able to outperform the often proposed naive allocation.

	Before transaction costs			After transaction costs			
	Mean	SD	SR	Mean	SD	SR	pc
$\omega_{\text{MVSF}}$	15.9	13.95	1.14	15.1	17.59	1.13	0.46
$\omega_{\text{LW}}$	14.1	13.03	1.06	13.8	13.8	1.00	0.24
$\omega_{\text{BRK}}$	14.5	16.47	0.88	14.0	15.90	0.88	0.22
Naive Portfolio	13.4	18.10	0.74	12.6	18.52	0.68	0.12
$\omega_{\text{GMVP}}$ w/o short	10.6	11.91	0.89	4.0	12.93	0.31	0.44

Table 5: Results of the bootstrapped empirical horse-race after transaction costs based on 1904 trading days and randomly choosing  $N = 100$  assets 100 times. *SR* denotes the (annualized) out-of-sample Sharpe ratio of the individual strategies. *Neg. W.* is the average of the absolute value of negative portfolio weights. *pc* is the portfolio concentration, computed as the average  $L_2$  norm of the portfolio weights. A description of the trading strategies is provided in Table 6.

## 7. CONCLUSIONS

The uncertain structure of real-world financial time series makes it hard to allocate wealth efficiently among many assets in order to maximize expected utility. We propose an optimization framework which incorporates model- and parameter uncertainty into the decision process and adjusts for transaction costs in a subsequent step. Our procedure allows to rely on the predictive performance of many different sources of information. We employ forecasting procedures based on high-frequency data and implement multivariate stochastic factor models. Within our approach we separate the task of predicting returns and reducing transaction costs, which is an advantage compared to widely used regularization approaches, as our framework addresses the objective function which we consider relevant for the investor: maximizing expected utility after adjusting for transaction costs. Based on analyzing portfolios generated based on NASDAQ data

from 2007 until 2016 we find that regularization is a key component in large dimensions and illustrate the importance of high-frequency data in the portfolio allocation context. We find that our model combination approach is not only successful in predicting returns but also outperforms a range of competitors in terms of portfolio performance.

## A. IMPLEMENTATION DETAILS

### A.1. BAYES-STEIN PROCEDURE

Jorion (1986) imposes regularization on the sample estimates  $\hat{\mu}$  and  $S$  of the expected mean and variance-covariance matrix to reduce estimation error. The estimate of the mean takes the form

$$\hat{\mu}_t^* = (1 - \psi)\hat{\mu}_t + \psi\hat{\mu}_t^g \quad (42)$$

where  $\hat{\mu}_t^g$  is defined as  $\hat{\mu}_t^g = \left( \frac{\hat{\mu}_t' \hat{\Sigma}_t^{-1} \iota}{\iota' \hat{\Sigma}_t^{-1} \iota} \right) \iota$ . The shrinkage parameter  $\psi \in (0, 1)$  is defined as

$$\psi = \frac{N + 2}{(N + 2) + h(\hat{\mu}_t - \hat{\mu}_t^g)' \hat{\Sigma}_t^{-1} (\hat{\mu}_t - \hat{\mu}_t^g)}. \quad (43)$$

In our application the estimation window  $h$  is chosen to be 250 trading days. The variance is estimated by shrinking  $S$  towards the variance of the minimum variance portfolio

$$\hat{\Sigma}_t^* = \left( 1 + \frac{1}{h + J_t} \right) S + \frac{J_t}{h(h + 1 + J_t)} \frac{\iota \iota'}{\iota' \hat{\Sigma}_t^{-1} \iota}. \quad (44)$$

The shrinkage factor  $J_t$  takes the form

$$J_t = \frac{N + 2}{(\hat{\mu}_t - \hat{\mu}_t^g)' \hat{\Sigma}_t^{-1} (\hat{\mu}_t - \hat{\mu}_t^g)} \quad (45)$$

The optimal portfolio choice rule  $\omega_{BS}$  is the solution to the optimization problem

$$\omega_{BS} = \arg \max_{\iota' \omega = 1} \omega' \hat{\mu}_t^* - \frac{2}{\gamma} \omega' \hat{\Sigma}_t^* \omega.$$

## B. PROOFS AND DERIVATIONS

*Proof of Proposition 1.* The allocation  $\omega$  is chosen to maximize the CE after transaction costs

$$\omega^* = \arg \max_{\omega \in \mathbb{R}^N: \iota' \omega = 1} \omega' \mu - \nu_t(\omega, \omega_{t^+}, \beta) - \frac{\gamma}{2} \omega' \Sigma \omega. \quad (46)$$

Equivalently, the optimization problem can be rewritten as choosing the rebalancing vector  $\Delta$ :

$$\omega^* = \omega_{t^+} + \arg \max_{\iota' \Delta = 0} (\omega_{t^+} + \Delta)' \mu - \beta \Delta' \Delta - \frac{\gamma}{2} (\omega_{t^+} + \Delta)' \Sigma (\omega_{t^+} + \Delta). \quad (47)$$

$$\Delta^* = \arg \max_{\iota' \Delta = 0} \underbrace{\omega_{t^+}' \mu - \frac{\gamma}{2} \omega_{t^+}' \Sigma \omega_{t^+} + \Delta' \mu - \beta \Delta' \Delta - \frac{\gamma}{2} \Delta' \Sigma \Delta - \gamma \Delta' \Sigma \omega_{t^+}}_{CE(\omega_{t^+})}. \quad (48)$$

$$\Delta^* = \arg \max_{\iota' \Delta = 0} \Delta' \mu - \frac{\gamma}{2} \Delta' \underbrace{\left( \frac{2}{\gamma} \beta I + \Sigma \right)}_{:=A} \Delta - \gamma \Delta' \Sigma \omega_{t^+}. \quad (49)$$

The first-order conditions take the form:

$$0 = \mu - \gamma A \Delta - \gamma \Sigma \omega_{t^+} - \lambda \iota \quad (50)$$

$$0 = \iota' \Delta \quad (51)$$

It follows that

$$\Delta = \frac{1}{\gamma} A^{-1} (\mu - \gamma \Sigma \omega_{t^+} - \lambda \iota) \quad (52)$$

Evaluating  $\iota' \Delta = 0$  with  $\Delta$  as above results in

$$\lambda = \frac{1}{\iota' A^{-1} \iota} \iota' A^{-1} \mu - \frac{\gamma}{\iota' A^{-1} \iota} \iota' A^{-1} \Sigma \omega_{t^+} \quad (53)$$

Plug-in gives:

$$\Delta_{\beta}^* = \frac{1}{\gamma} A^{-1} \mu - A^{-1} \Sigma \omega_{t^+} - \frac{1}{\gamma \iota' A^{-1} \iota} A^{-1} \iota' A^{-1} \mu \iota + \frac{1}{\iota' A^{-1} \iota} A^{-1} \iota' A^{-1} \Sigma \omega_{t^+} \iota \quad (54)$$

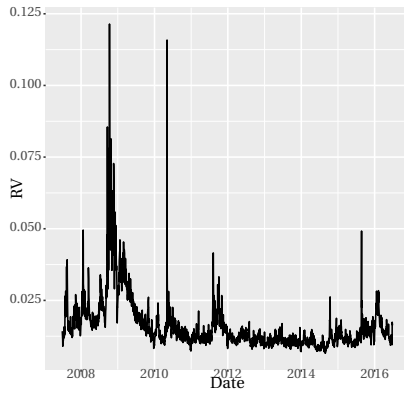
$$= \frac{1}{\gamma} \left( A^{-1} - \frac{1}{\iota' A^{-1} \iota} A^{-1} \iota \iota' A^{-1} \right) \mu - A^{-1} \Sigma \omega_{t^+} + \frac{1}{\iota' A^{-1} \iota} A^{-1} \iota' A^{-1} \Sigma \omega_{t^+} \iota \quad (55)$$

□

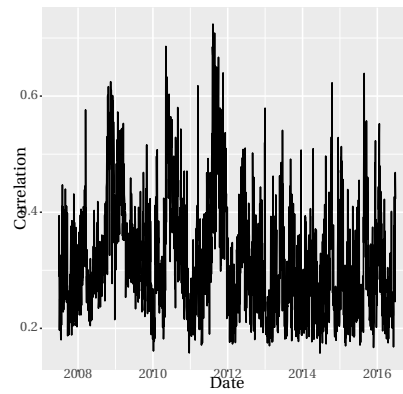
## C. FIGURES AND TABLES

Table 6: Implemented portfolio allocation strategies. Strategies abbreviated with the suffix unrestr denote the solution of maximization problem (EU) without incorporating a penalty on turnover by setting  $v_t \equiv 0$ . This corresponds to choosing the maximum expected utility allocation irrespective of the costs of realization. The competitors  $\omega_{\text{Naive}}$ ,  $\omega_{\text{KaZ}}$  and  $\omega_{\text{JB}}$  denote the equal weighted portfolio, the procedure described by Kan and Zhou (2007) and the Bayes-Stein approach proposed by Jorion (1986), respectively. Details regarding the implementation of  $\omega_{\text{KaZ}}$  and  $\omega_{\text{JB}}$  are provided in the Appendix.

Name	Description	Abbreviation
Stochastic Volatility	Bayesian Multivariate Stochastic Volatility Factor Model (See section 4.2)	$\omega_{\text{MVSF}}$ $\omega_{\text{MVSF, unrestr}}$
Ledoit and Wolf (2004)	Regularized Variance-Covariance Matrix (See section 4.3). Return predictions are computed using $r_{t+1} \sim N(0, \hat{\Sigma}_t^{LW})$	$\omega_{\text{LW}}$ $\omega_{\text{LW, unrestr}}$
Sample Estimate	Predictive return distribution computed under assumption of normal distributed returns with variance covariance matrix equal to the sample estimator $S$	$\omega_S$ $\omega_{S, unrestr}$
Normal-Wishart BRK	Block-Realized-Kernel estimator based on 6 groups used as input to compute predictive return distribution under the Normal-Wishart model (See section 4.1)	$\omega_{\text{BRK, NW}}$ $\omega_{\text{BRK, NW, unrestr}}$
Hautsch et al. (2012)	Block-Realized-Kernel estimator based on 6 groups used as input to compute predictive return distribution under the assumption of normality	$\omega_{\text{BRK}}$ $\omega_{\text{BRK, unrestr}}$
Geweke (2011)	Mixture based on optimal prediction pooling (See Section 3)	$\omega_{\text{Mixing}}$ $\omega_{\text{PP, unrestr}}$
Naive Portfolio	Equal weight portfolio ( $\omega = 1/N$ )	$\omega_{\text{Naive}}$
Jorion (1986) Bayes-Stein	Parameter smoothing by Bayesian shrinkage (See Section A.1)	$\omega_{\text{JB}}$
GMVP	Global Minimum Variance Portfolio $\omega_{\text{GMVP}} := \frac{S^{-1} \mathbf{1}}{\mathbf{1}' S^{-1} \mathbf{1}}$ . Additionally, GMVP with sort-selling constraint Jagannathan and Ma (2003)	$\omega_{\text{GMVP}}$ $\omega_{\text{GMVP w/o short}}$



(a) Realized Volatilities



(b) Average Correlations

Figure 1: Cross-sectional averages of daily realized volatilities and absolute correlation estimates based on intra-daily data for  $N = 100$  assets. Realized volatilities are computed using the methodology proposed by Zhang (2011). The second panel denotes the mean of the absolute entries in the estimated BRK-matrices as described in Section 2.

Sector	Number of Ticker
Health Care	9
Consumer Durables	2
Capital Goods	7
Finance	12
Consumer Services	21
Energy	11
Public Utilities	11
Transportation	3
Technology	14
Basic Industries	5
Consumer Non-Durables	3
Miscellaneous	2

Table 7: Industry classification (according to NASDAQ) of the  $N = 100$  assets chosen from the asset universe used to create the results in Sections 2 and 6.



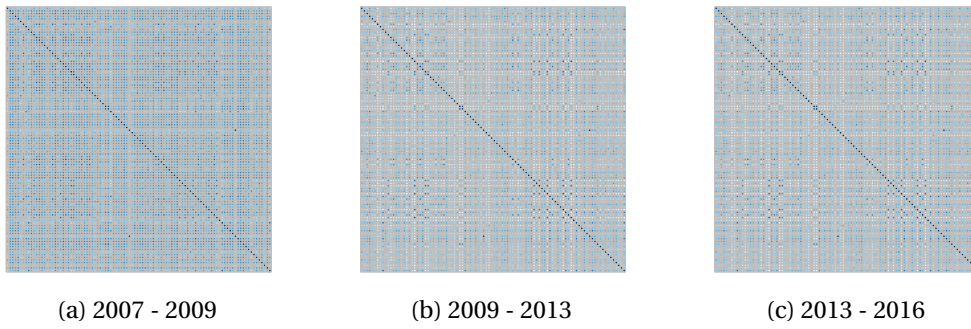


Figure 2: Empirical correlation matrices for the  $N = 100$  assets based on daily observations. Brighter values indicate lower correlations among the assets. During the sample period only positive correlations are measured. The darker, the higher the correlations. The first figure shows estimates before and during the financial crisis (June 2007 until June 2009), Panel b) denotes the time horizon between June 2009 and June 2013 and Panel c) visualizes the estimates based on the observations between June 2013 and June 2016. Correlations are time-varying and especially large during the crisis in our sample.

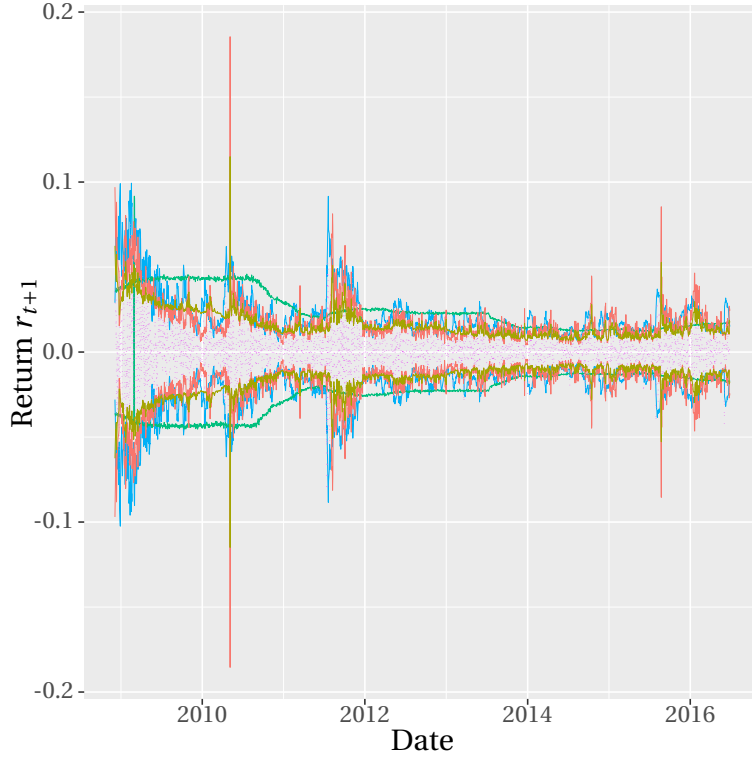


Figure 3: Illustrative results to compare the predictive performance of the stochastic volatility factor model with 2 factors, the block realized-kernel estimates and the Ledoit-Wolf regularization. For each of the three models a sample of the predicted portfolio return of the  $1/N$  allocation is generated by first sampling from the predictive return distribution  $r_{t+1}^{(m)}$  and then evaluating the vector  $\frac{1}{N}l'r_{t+1}^{(m)}$ . The blue line indicates the 0.05 and 0.95 quantiles of the prediction according to the SV Model, the green line shows the corresponding quantiles of the Ledoit-Wolf approach and the yellow line corresponds to the BRK estimates. The red line corresponds to the forecasts based on our Normal-Wishart model. The purple dots indicate the observed return at time  $t + 1$ .

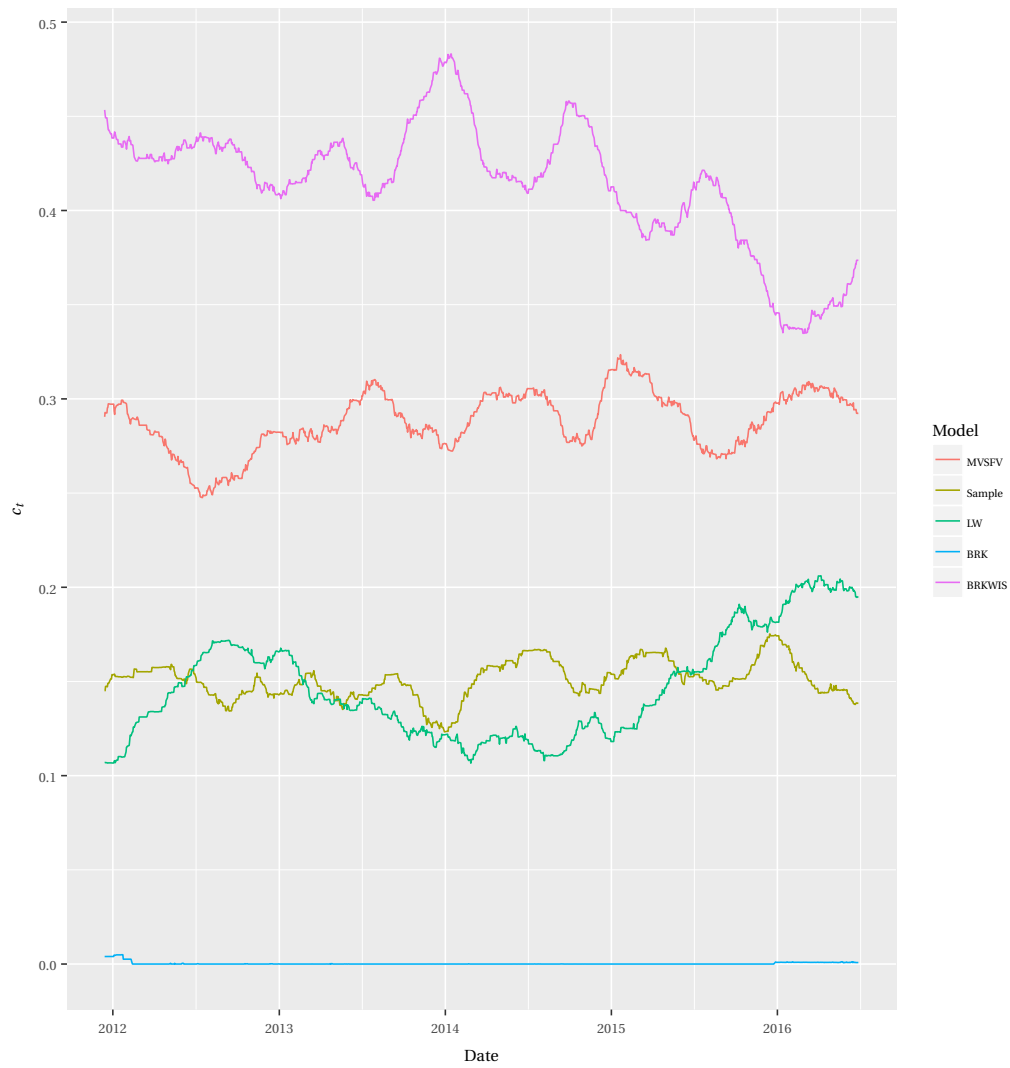


Figure 4: Optimal weighting schemes  $c_t$  based on Geweke and Amisano (2011).

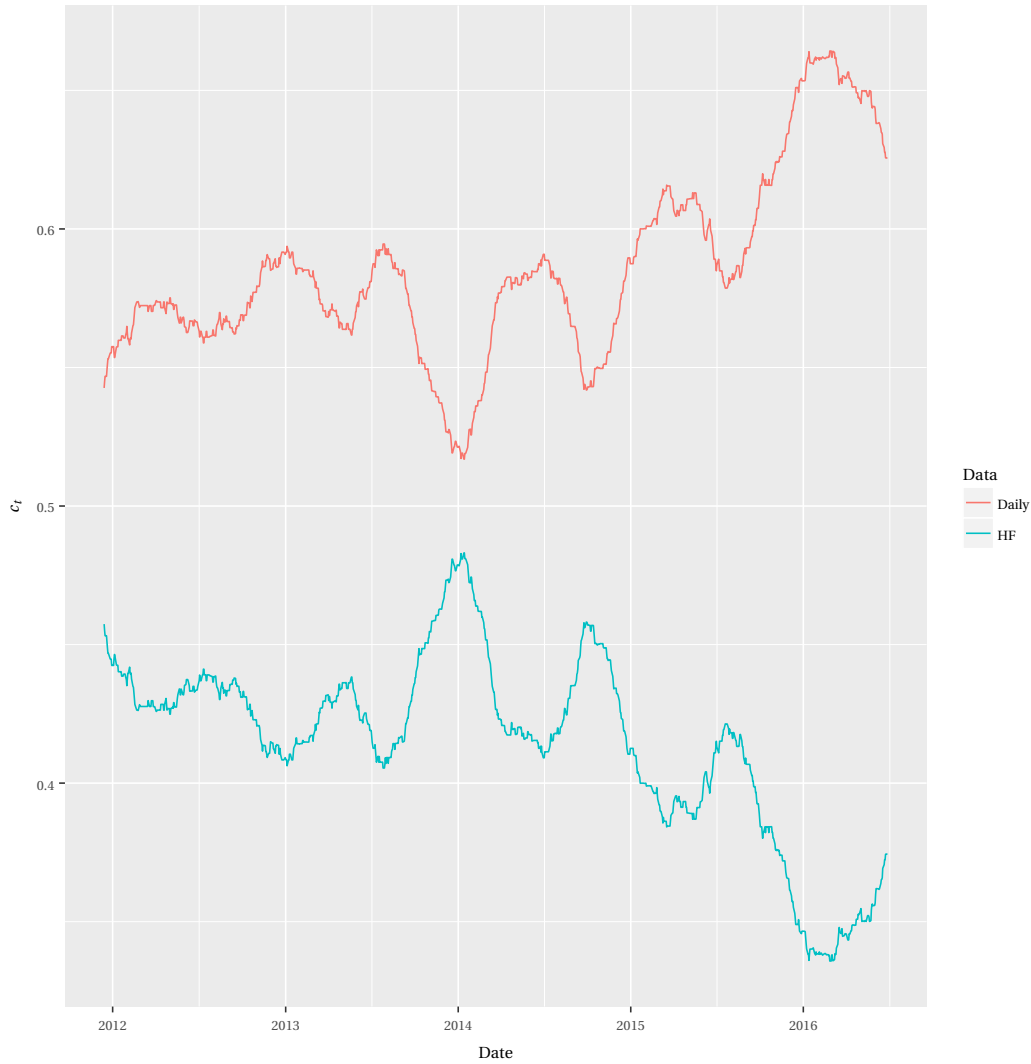
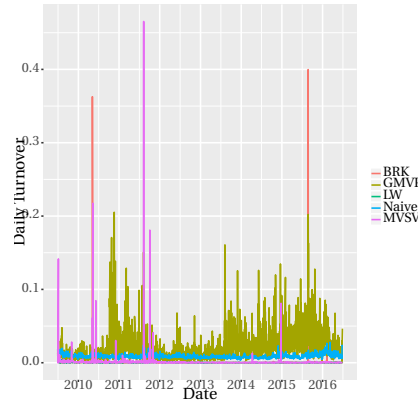
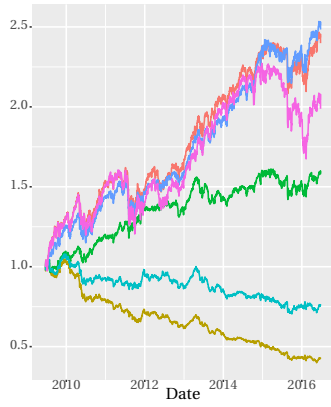


Figure 5: Optimal weighting schemes conditional on data input. Time-series of model combination weights based on the predictive performance. HF corresponds to the cumulative probability of the BRK modes, daily corresponds to the sum of the combination weights resulting from the Ledoit-Wolf approach, the sample covariance estimator and the stochastic volatility factor model.



(a) Cumulative wealth for the distinct portfolio allocation strategies (starting value of 1)

(b) Daily turnover measured as  $L_1$  norm of the distinct portfolio allocation strategies.

Figure 6: Summary statistics of the performance of the distinct allocation strategies based on  $N = 100$  assets. The strategies implemented are described in Table 6. Plotted are only the results of optimization strategies which take into consideration  $v(\omega)$  (unrestricted strategies exhibit much more turnover).

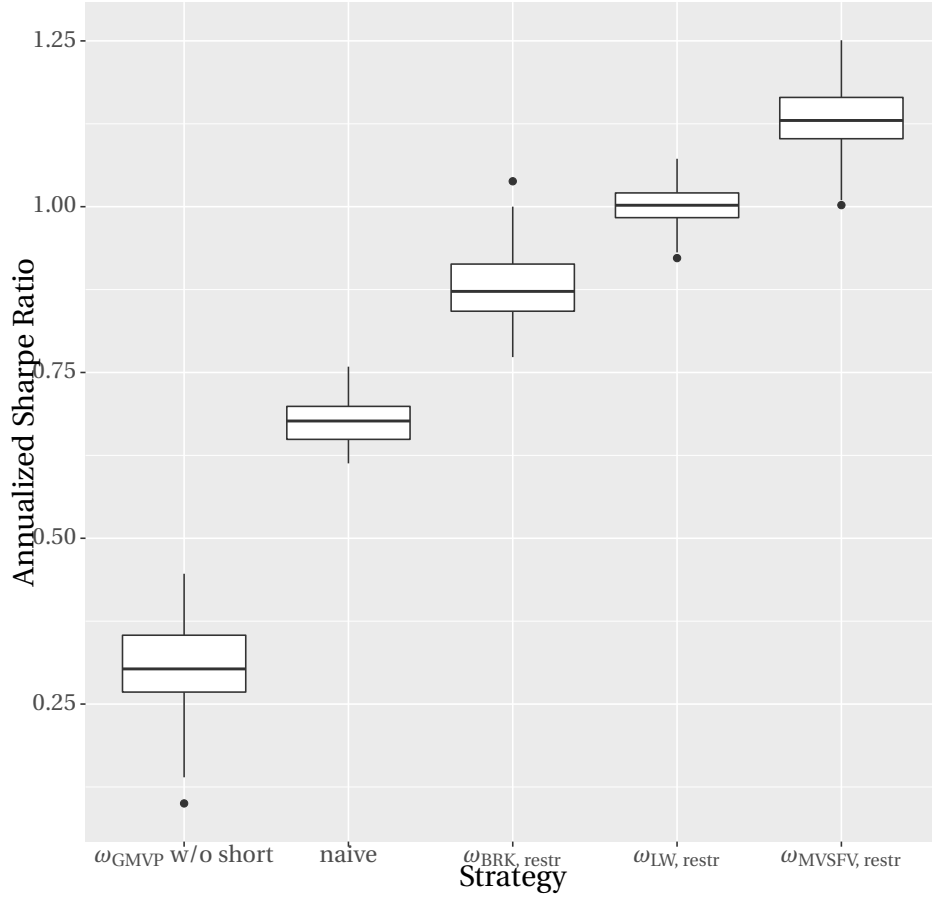


Figure 7: Annualized Sharpe ratios of the distinct portfolio allocation strategies based on portfolio bootstrapping. We compute the optimization procedure by picking  $N = 100$  assets randomly for  $s = 75$  times. For each sample  $s$  a complete horse-race is performed as described in Section 6. The box plots denote the out-of-sample Sharpe ratios after adjusting for transaction costs. The corresponding strategies are listed in Table 6.

## REFERENCES

- Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357.
- Aït-Sahalia, Y. and Xiu, D. (2015). Principal component estimation of a large covariance matrix with high-frequency data.
- Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- Anderson, E. W. and Cheng, A.-R. M. (2016). Robust Bayesian Portfolio Choices. *Review of Financial Studies*, page hhw001.
- Andersson, J. (2001). On the normal inverse gaussian stochastic volatility model. *Journal of Business & Economic Statistics*, 19(1):44–54.
- Arnott, R. D. and Wagner, W. H. (1990). The measurement and control of trading costs. *Financial Analysts Journal*, 46(6):73–80.
- Bannouh, K., Martens, M., Oomen, R. C. A., and van Dijk, D. J. C. (2012). Realized mixed-frequency factor models for vast dimensional covariance estimation. *ERIM Report Series Reference No. ERS-2012-017-F&A*.
- Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, 24(1):1–13.
- Barndorff-Nielsen, O. E. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of econometrics*, 162(2):149–169.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.
- Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Or*, pages 451–468.
- Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Dynamic Econometric Modeling and Forecasting*, 177(2):213–232.
- Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.
- Brandt, M. W. (2010). Portfolio choice problems. In Ait-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics*, volume 1, pages 269–336. North-Holland Publ. Co., Amsterdam.
- Brodie, J., Daubechies, I., de Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272.
- Brown, S. J. (1976). *Optimal portfolio choice under uncertainty: a Bayesian approach*. PhD Thesis, University of Chicago, Chicago.
- Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.
- Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of econometrics*, 134(2):341–371.
- Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, pages 135–155.



- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.
- Durham, G. and Geweke, J. (2014). Improving asset price prediction when all models are false. *Journal of Financial Econometrics*, 12(2):278–306.
- Fan, J., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):412–428.
- Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics*, 67(3):473–509.
- Forsberg, L. and Bollerslev, T. (2002). Bridging the gap between the distribution of realized (ecu) volatility and arch modelling (of the euro): The garch-nig model. *Journal of Applied Econometrics*, 17(5):535–548.
- Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81.
- Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340.
- Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of econometrics*, 164(1):130–141.
- Greyserman, A., Jones, D. H., and Strawderman, W. E. (2006). Portfolio selection using hierarchical Bayesian analysis and MCMC methods. *Journal of Banking & Finance*, 30(2):669–678.
- Halbleib, R. and Voev, V. (2016). Forecasting covariance matrices: A mixed approach. *Journal of Financial Econometrics*, 14(2):383–417.
- Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *Review of Financial Studies*, 19(1):237–271.
- Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of applied econometrics*, 27(6):877–906.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., and Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do High-Frequency Data Improve High-Dimensional Portfolio Allocations? *Journal of applied econometrics*, 30(2):263–290.
- Hautsch, N., Kyj, L. M., and Oomen, R. C. A. (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of applied econometrics*, 27(4):625–645.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401.
- Idzorek, T. M. (2002). A step-by-step guide to the black-litterman model. *Forecasting expected returns in the financial markets*, page 17.
- Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87.
- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684.
- Jensen, M. B. and Lunde, A. (2001). The nigs&arch model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *The Econometrics Journal*, 4(2):319–342.
- Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics*, 11(2):335–369.
- Jobson, J. D., Korkie, B., and Vinod, R. (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the Business and Economic Statistics Section*, pages 279–284.
- Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644.
- Jondeau, E. and Rockinger, M. (2012). On the importance of time variability in higher moments for asset allocation. *Journal of Financial Econometrics*, 10(1):84–123.
- Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(03):279–292.

- Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(03):621–656.
- Kastner, G. (2016). factorstochvol: Bayesian estimation of (sparse) latent factor stochastic volatility models.
- Laloux, L., Cizeau, P., Bouchaud, J.-P., and Potters, M. (1999). Noise dressing of financial correlation matrices. *Physical review letters*, 83(7):1467.
- Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.
- Liu, C., Xia, N., and Yu, J. (2016). Shrinkage estimation of covariance matrix for portfolio choice with high frequency data. *arXiv preprint arXiv:1611.06753*.
- Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data? *Journal of applied econometrics*, 24(4):560–582.
- Lopes, H. F., Kastner, G., and Frühwirth-Schnatter, S. (2016). Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.
- Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of applied econometrics*, 27(6):907–933.
- Perold, A. F. (1984). Large-scale portfolio optimization. *Management Science*, 30(10):1143–1160.
- Pettenuzzo, D. and Ravazzolo, F. (2016). Optimal Portfolio Choice Under Decision-Based Model Combinations. *Journal of applied econometrics*, pages n/a–n/a.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68.
- Tu, J. and Zhou, G. (2010). Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, 45(04):959–986.
- Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

- Uppal, R. and Wang, T. (2003). Model misspecification and underdiversification. *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.
- Yoshimoto, A. (1996). The mean-variance approach to portfolio optimization subject to transaction costs. *Journal of the Operations Research Society of Japan*, 39(1):99–117.
- Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of econometrics*, 160(1):33–47.