Efficiency in high-dimensional portfolio allocation: comparison using high-frequency data

M. Chudý\textsuperscript{a,*}, N. Hautsch\textsuperscript{a}

\textsuperscript{a}Department of Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

Abstract

Our goal is to assess the empirical performance of the local method of moments estimator (LMME) of integrated covolatility from noisy and asynchronous high-frequency data in a high-dimensional framework. We propose necessary adjustment for the LMME to accommodate for the curse of dimensionality. As these adjustments may lead to loss of efficiency, we use Kulback-Leibler distance to measure the discrepancy between original and adjusted LMME. A global minimum variance portfolio exercise provides basis for our out-of-sample comparison of LMME with selected competitors. Our results show that exploiting full correlation structure of the assets leads to lower portfolio volatility and better diversification but requires more frequent re-balancing of the assets. The benefits of LMME against its competitors are reported in terms of economic utility for a risk-averse investor.

Keywords: local method of moments, cholesky factorization, high-dimensional portfolio, micro-structure noise, asynchronous high-frequency data, graphical lasso, hierarchical correlation clustering, smoothing, Kulback-Leibler distance

1. Introduction

One of the key questions of financial risk management is optimal portfolio allocation. With the availability of big high-frequency (HF) asset price data and simultaneous rise of computational power, the focus of both research and practice has turned to vast portfolios counting hundreds of highly liquid assets. From side of econometrics, optimal asset
allocation translates into finding a feasible and efficient covariance estimator from the observed prices. Ex-ante forecasts of the portfolio covariance matrices are particularly useful for optimal portfolio allocation as they reduce the future portfolio volatility, i.e., the riskiness of the investment. On the other hand, in real life, transaction costs must be taken into account for evaluation of any investment strategy. Hautsch et al. (2015) showed that even for high-dimensional portfolios, HF covolatility estimators generally dominate in terms of portfolio volatility. But even when smoothed over time, HF estimators lead to higher transaction costs, because they require frequent re-balancing of the portfolio. Up to certain point, these additional transaction costs are still worth for an investor to switch from low-frequency to HF estimators. Hautsch et al. (2015) showed that this may be particularly true during periods such as crises, when new information arrive every second. Still, even in such turbulent periods, it is very important that the predicted portfolio weights show certain amount of stability over the investment horizon.

Usual problems with HF covolatility estimators arise from various sources of pollution which is always present in the observed data. The micro-structure noise is a typical example of such pollution. But as the dimension of the portfolio grows, other issues like asynchronicity and differences in liquidity of portfolio constituents become serious. Estimators which can deal with all sorts of pollution and problems related to the curse of dimensionality exist. However, there are only few of them, which exploit other HF assets characteristics such as liquidity. The regularization and blocking (RnB) estimator (Hautsch et al., 2012) can deal with both micro-structure noise and asynchronicity. The estimator uses refresh-time sampling on assets divided into groups according to their liquidity. Without the blocking step, refresh-time sampling can lead to data loss fraction of 80%. On each block, the multivariate realized kernel estimator (MRK, see Barndorff-Nielsen et al., 2011) is computed and the blocks are “put” together to form the final estimators. Boudt et al. (2017) pushed this idea to the limit introducing the CHOLCOV estimator which estimates the covariance matrix element-by-element by exploiting the cholesky factorization. The discrimination between assets based on their liquidity has clear benefits for portfolios, where both highly liquid and illiquid assets are present. On the opposite side of the element-by-element estimation (see also Aït-Sahalia et al., 2010; Fan et al., 2012; Lunde et al., 2016, for more estimators of this type) are estimators which exploits information about the full dependency structure among portfolio constituents such as LMME of Bibinger et al. (2014). A natural question that arises from availability of such estimators is whether an investor could benefit from using an efficient HF estimator that exploits the full correlation structure rather than just “any” HF estimator. The question is of empirical

---

1 E.g. in our SP 500 portfolio, one of most liquid assets AAPL has typically 40000 intra-day observations (after revision) whereas least liquid assets such as XRX may have only 100.
nature and we will try to find the answer in a one-step-ahead ex-ante forecasting comparison. Some past empirical studies have shown that in real-life situations, even simple procedures can dominate over sophisticated ones, particularly in out-of-sample forecasting. Therefore, we conduct an empirical competition of selected HF estimators based on a global minimum portfolio exercise similar to Hautsch et al. (2015). Note that compared to the latter paper we 1. address a different research question, 2. use different pool of estimators and 3. use different data source².

The goal of this paper is to empirically compare the efficient LMME proposed by Bibinger et al. (2014) which exploits the full correlation structure against the non-efficient element-by-element CHOLCOV estimator of Boudt et al. (2017), which exploits the hierarchical ordering of elements of the triangular matrices in cholesky factorization thus effectively uses available observations for each bivariate covariance. The common basis for our comparison is provided by the local spectral estimator proposed by Bibinger and Reiss (2014). The latter estimator is used as a pilot to obtain the efficient LMME and at the same time as the bivariate estimator of each element contained in CHOLCOV (see the implementation details in the next section). We will refer to the estimator of Bibinger and Reiss (2014) as PILOT further on. Each of the three estimators has its advantages and disadvantages. The PILOT estimator is computationally feasible (both in terms of computational memory & speed) even when the dimension of portfolio is large. It can easily deal with asynchronicity thanks to the spectral transformation and does not require any synchronization which is time-consuming when portfolio counts hundreds of assets. The CHOLCOV estimator is guaranteed positive semi-definite. This property, however, is of little use for applications such as in this paper, where strict definiteness and well-conditioned matrix estimator are required. On the other hand, the CHOLCOV estimator benefits from cholesky factorization that allows using a bivariate estimator for each covariance element. Its major drawback is that the computation requires repeated synchronization of growing number of assets which must be done sequentially due to the hierarchical factorization. This leads to a very long computation (for the SP 500 constituents the computation of the estimator took 2 hours per trading day) and limits the practical availability. Finally, LMME is the only efficient of the three estimators thanks to the optimal weighting matrices

\[ W_{jk} = I^{-1}_k I_{jk}, \]  

for each spectral frequency \( j = 1, \ldots, J \), and on each intra-day interval \( k = 1, \ldots, K \). The

---

²Hautsch et al. (2015) used the high-frequency midquote data from TAQ database, whereas we use so called tick-level data (i.e. sampled with highest possible frequency) from LOBSTER: Limit Order Book System - The Efficient Reconstructor.
major limitation is however related to the Fisher-type matrices, which can get infeasibly large even for extraordinary computation facilities such as the 256 GB nodes available on Vienna Scientific Cluster. For the SP 500 constituents, these matrices would require allocation of approx. $J \times K \times 123$ GB, which is, up to our knowledge, beyond current limits of most software packages including Matlab. Additionally, a decomposition of the weights is not possible due to the inverse of the matrix $I_k = \sum_j I_{jk}$ in (1). Practically, the estimator becomes infeasible on any ordinary machine for $m \geq 50$. To ensure feasibility, we propose adjustments of the LMME estimator. Besides the practical application requires all estimators to be stable, positive definite and well-conditioned. However, due to the curse of dimensionality, the latter properties are not guaranteed by any of the estimators above. In fact the PILOT and LMME are not even guaranteed to be positive semi-definite. Therefore, we employ regularization and smoothing steps for all three estimators to strike some balance between the responsiveness and stability of each estimator (see Bai and Shuzhong, 2011, for discussion on various regularization methods).

In order to facilitate the use of LMME in high-dimensional portfolios we have to reduce its computational memory requirement. Clearly, one could use similar approach as Hautsch et al. (2012) and group the assets based on their liquidity. By limiting the size of each group, one could then easily estimate LMME group-wise. However, as LMME benefits from exploit correlation structure among multiple assets irrespective of their liquidity, it is appropriate to identify the groups based on their correlation matrix. This can be done using hierarchical correlation clustering (see Hastie et al., 2009).

The group-wise estimation does not guarantee that the compound estimator has the required properties, after all pieces are back together. Therefore we look for a machine-learning technique, that would conveniently combine the blocking and regularization step. Only recently it has been shown that graphical LASSO (GLASSO) proposed by Friedman et al. (2008) performs exactly these two steps. First it identifies the clusters based on a correlation matrix and then, using penalized likelihood, estimates the inverse covariance matrix (i.e. the precision matrix) for each cluster separately, forcing the off-block-diagonal elements to 0. Tan et al. (2015) use this surprising finding to enhance the performance of GLASSO by different choice of linkage functions (see Table 1).

The block-wise estimation of LMME would be efficient only if the true covolatility were block-diagonal. Typically, the covolatility of assets returns does not have such structure and the approach certainly leads to loss of the information. In order to get idea about the amount of information that is lost, i.e., how far we got from the original LMME we use the Kulback-Leibler (K-L) measure as proposed by Tumminello et al. (2007, 2010).

Finally, we evaluate the benefits of the adjusted LMME against PILOT and CHOLCOV based on various portfolio characteristics computed under the global minimum variance (GMV) framework. Following Hautsch et al. (2012) we also look at economic utility of a
risk averse investor gained by switching from PILOT/CHOLCOV to LMME.

The rest of the paper is organized as follows: in section 2 we give a brief summary of the estimators mentioned above, section 3 deals with the clustering and regularization adjustments to LMME, section 4 compares the original with adjusted LMME from informational loss perspective, section 5 presents empirical results and section 6 concludes.

2. The estimators of integrated covolatility under noise and asynchronicity

We assume that the latent (true) $m$-dimensional price process follows a continuous martingale

$$y_t^* = y_0^* + \int_0^t \Sigma^{1/2}(s) dB_s, \quad t \in [0,1],$$

with $B_s$ denoting an $m$-dimensional standard Brownian motion process. The asynchronous price observations are polluted by micro-structure noise:

$$y_{t,i,l} = y_{t,i}^* + \epsilon_{i,l}, \quad \epsilon_{i,l} \sim N(0, \eta^2_l), \quad i = 1, \ldots, n_l, \quad l = 1, \ldots, m.$$  

We are estimating integrated covolatility

$$\int_0^1 \Sigma(s) \, ds.$$  

2.1. PILOT

The local spectral estimator proposed by Bibinger and Reiss (2014) is obtained from log-returns $x_{t,i,l} = \log y_{t,i,l} - \log y_{t-1,i,l}, \quad i = 1, \ldots, n_l, \quad l = 1, \ldots, m$. First, the trading day is partitioned into $K$ intervals of length $h = 1/K$. On each intra-day interval $[kh, (k+1)h], \quad k = 1, \ldots, K - 1$ the spot covolatility $\Sigma_k(s)$ is assumed constant hence the integrated covolatility on each interval is obtained as $h \Sigma_k$. The PILOT estimator is just a sum of these local estimators

$$\text{PILOT} = \frac{1}{K} \sum_{k=1}^K \hat{\Sigma}_k,$$  

which can be computed in parallel on each block as follows:

(i) For each asset $l = 1, \ldots, m$, compute the squared duration\(^3\)

$$\nu_l^2 = \sum_{t_{i,l} \in [kh,(k+1)h]} (\Delta t_{i,l})^2.$$  

\(^3\)Also called “quadratic variation of times” in Bibinger et al. (2014).
(ii) For each asset \( l = 1, \ldots, m \), compute a series of spectral averages

\[
s_{jk,l} = \sqrt{2/h} \sum_{t_{i,l} \in [kh, (k+1)h]} x_{t_{i,l}} \sin \left( j\pi h^{-1}(\bar{t}_{i,l} - kh) \right), \quad j = 1, \ldots, J, \tag{7}
\]

with \( \bar{t}_{i,l} = \frac{t_{i,l} + t_{i,l-1}}{2} \). These spectral averages are orthogonal over dimension \( j \) and form an analogy to principal components in spectral domain. Hence they carry a maximal information about the variability attributed to frequency \( j \). The spectral cut-off \( J \) should be possibly small (e.g. in our case \( \leq 5 \)). Additionally, under general assumptions on the underlying price process, the averages have the property: \( s_{jk} \sim \mathcal{N}_m(0, C_{jk}) \), with \( C_{jk} = \Sigma_k + H_{jk} \) where \( \Sigma_k \) is \((m \times m)\) spot covolatility matrix to be estimated. In the bias term \( H_{jk} = (j\pi/h)^2 \text{diag} \left\{ \nu_l^2 \eta_l^2 \right\} \) is the scaled noise level and stems both from irregular arrival times and micro-structure noise variance \( \eta_l^2 \). The latter can be estimated from the observed log-returns as

\[
\hat{\eta}_l^2 = \frac{1}{2(n_l - 1)} \sum_{i=1}^{n_l} (x_{i,l})^2. \tag{8}
\]

(iii) Compute the unbiased spot covolatility matrix estimator

\[
\hat{\Sigma}_k = \frac{1}{J} \sum_{j=1}^{J} \left( s_{jk}^{\otimes 2} - H_{jk} \right). \tag{9}
\]

2.2. LMME

Bibinger et al. (2014) propose, in analogy with the classical GMM approach by Hansen (1963), a localized method of moments estimator LMME to improve the efficiency of PILOT by employing weights \( W_{jk} \) for each frequency \( j = 1, \ldots, J \). The weights depend on the Fisher-type \((m^2 \times m^2)\) information matrices

\[
I_{jk} = C_{jk}^{-1} \otimes C_{jk}^{-1} \quad \text{and} \quad I_k = \sum_{j=1}^{J} I_{jk}. \tag{10}
\]

So the (vectorized) weighted analogue of the of the spot covolatility estimator from (9) is

\[
\text{vec}\hat{\Sigma}_k^{\text{adap}} = \sum_{j=1}^{J} W_{jk} \text{vec} \left( s_{jk}^{\otimes 2} - H_{jk} \right). \tag{11}
\]

Hence the LMME is obtained as:

\[
\text{LMME} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Sigma}_k^{\text{adap}}. \tag{12}
\]
2.3. CHOLCOV

The estimator of Boudt et al. (2017) exploits the hierarchical structure implied by the
cholesky factorization in order to effectively use most of the available observations. The
idea is to apply cholesky factorization to the spot covolatility on each intra-day interval

$$\Sigma_k = BVB^\top, \quad (13)$$

where

$$B = \begin{bmatrix} 1 & 0 & \ldots & 0 \\
1 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \ldots & 0 \\b_{m1} & b_{m2} & \ldots & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_{11} & 0 & \ldots & 0 \\
0 & v_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & v_{mm} \end{bmatrix}. \quad (14)$$

The elements $b_{ij}$ and $v_{ii}, i, j = 1, \ldots, m$ and $j < i$, can be estimated by sequential
algorithm (see page 351 in Boudt et al., 2017) and one can chose any bivariate esti-
mator for the elements of CHOLCOV. We chose the bivariate local method of Bibinger
and Reiss (2014). Again the computation proceeds in parallel on the intra-day intervals $[kh, (k+1)h], k = 1, \ldots, K - 1$ with following steps:

(i) sort the stocks according to decreasing liquidity. As criterion for liquidity, we use
the scaled squared duration\footnote{Boudt et al. (2017) suggest using only squared duration without division by number of observations. Yet, at least in this case, it lead to quite disproportional ordering, when some less liquid assets got higher priority which lead to unnecessary waste of observations.} \(\nu_t^2/n_l\) as defined in (6). Now the asset log-prices sorted according to liquidity are $y_{t,(l)}, l = 1, \ldots, m$.

(ii) set $v_{11} = \hat{\Sigma}_{11,k}$, the univariate analogue of (9) for most liquid asset $y_{t,(1)}$.

(iii) for $i = 2, \ldots, m$,

- synchronize the prices for assets $(l) = (1), \ldots, (i)$ and compute the refresh-
time synchronized log-returns $x_{\tau,(l)}, \tau = 1, \ldots, n_i$.
- set $f_1 = x_{(1)}$.
- for $j = 1, \ldots, i-1$ compute

$$b_{ij} = \hat{\Sigma}_{21,k}/\hat{\Sigma}_{22,k}\quad (15)$$

$$f_{j+1} = x_{(j+1)} - f_jb_{ij} - \ldots - f_1b_{i1}, \quad (16)$$

where $\hat{\Sigma}_k$ is the bivariate analogue of (9) for $(x_{(i)}, f_j)$. 

set $v_{ii} = \hat{\Sigma}_{11,k}$, which is the univariate analogue of (9) for $f_i$. 

(iv) the spot estimator on the intra-day interval is

$$\Sigma_{chol}^k = BVB^\top.$$

Finally, the estimator is obtained as

$$\text{CHOLCOV} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Sigma}_{chol}^k. \quad (18)$$

### 3. Grouping & Regularization for LMME

An established method to identify groups of assets such that the dependence is strong within each group but small among the groups is the hierarchical correlation clustering (HCC, see Hastie et al., 2009). In HCC the adjacency matrix, which represents the distances between pairs of assets is defined as $(1 - |R|)$, where $R \in \mathbb{R}^{m \times m}$ is the correlation matrix of the whole portfolio. At each step the two closest assets/clusters are merged into a new cluster until all assets belong to a single cluster. The distance between two clusters is determined by the linkage function (see Table 1). Once the number of clusters has reached preassigned limit, the procedure can be stopped. We can then compute the LMME estimator on each cluster. The question is, what should be done with the missing off-cluster elements of the matrix $\hat{\Sigma}_{adap}^k$. Setting them equal 0 would be too restrictive, therefore, we simply substitute these missing elements by respective estimates from the PILOT. A related problem is the conditionality of the final estimator. LMME as suggested by Bibinger et al. (2014) is not necessarily positive definite, but if the spectral cut-off $J$ and the number of intra-day intervals $K$ are small\(^5\), the LMME usually is positive definite. However, as the curse of dimensionality, the estimator becomes extremely ill-conditioned. Later on, we measure the forecasting performance in terms of volatility-minimizing portfolio weights:

$$\omega^* = \frac{\Xi_t}{t^\top \Xi_t}, \quad (19)$$

where $\Xi = \Sigma^{-1}$, is the precision matrix. If the covolatility estimator is ill-conditioned, a small change in any of its elements might result in a large change in $\Xi_t$. Beside the numerical instability, it also leads to large portfolio turnover thus transaction costs. To achieve

\(^5\)Proportionality parameters which can help to find right values of $J$ and $K$ can be found in Bibinger et al. (2017).
more stability in the allocation performance, we employ a regularization in the estimation process. Some popular regularization methods include shrinkage, latent factor models, random matrix theory and Bayes method. Hautsch et al. (2015) compares the first three methods but finds little evidence for superiority of one or the others. All these methods focus on regularization of the covariance estimator itself. An alternative approach would be to focus on the precision matrix $\Xi$, whose elements are proportional to the conditional correlations\(^6\). A technique that exploits the sparsity of the precision matrix is for instance the graphical lasso (GLASSO) (see e.g. Friedman et al., 2008). The GLASSO estimator of $\Xi$ is:

$$\hat{\Xi}_\lambda = \arg\min_{\Xi > 0} \left( \text{Tr}(\Xi \hat{R}) - \log \det(\Xi) + \| L \circ (\Xi - \text{diag}(\Xi)) \|_1 \right) , \quad (20)$$

for a given tuning parameter matrix $L_{ij} = \lambda > 0$ and the sample correlation matrix estimator $\hat{R}$. The symbol $\circ$ denotes entrywise product.

Furthermore, the GLASSO can be seen as a two step procedure: first the single-linkage HCC is performed on the correlation matrix $R$ to identify the clusters. Second, the regularization is performed by solving (20) with following penalty parameter restriction:

$$L_{ij} = \begin{cases} \lambda, & \text{should assets } i \text{ and } j \text{ be in the same cluster,} \\ \infty, & \text{otherwise.} \end{cases}$$

The value of $\lambda$ plays an important role in the HCC step too. It determines the height of the cluster tree (dendrogram) at which the clustering is stopped. Tan et al. (2015) suggest taking advantage of this decomposition of GLASSO and alternate the first step by using different linkage function. They argue that the single linkage results in so called trailing clusters, i.e. situation where majority of assets belongs to one huge cluster while the rest of clusters is univariate. Therefore, they suggest using average or complete linkage functions (see Table 1) to enforce more balanced among the clusters. We exploit this proposal of Tan et al. (2015) in the LMME estimation. The final estimator is obtained by the following algorithm applied locally on each intra-day interval $k = 1, \ldots, K$:

(i) Compute the local spot covolatility estimator $\hat{\Sigma}_k$ as in (9) and obtain the respective correlation estimator $\hat{R}_k = D^{-1/2}\hat{\Sigma}_k D^{-1/2}$, where $D = \text{diag}(\hat{\Sigma}_k)$.

(ii) Apply HCC with average linkage function and adjacency matrix $(1 - |\hat{R}_k|)$.

\(^6\)Correlations between returns for pair of assets given the returns of all other assets.
(iii) Search the cluster tree top to bottom (starting with 1 cluster). Cut the tree at the highest point where each cluster \( C_c \subseteq \{1, \ldots, m\}, c = 1, \ldots, M \) contains less than 50 assets. Save the high \( \lambda \) of the tree at where it is cut.

(iv) For each cluster obtain the GLASSO estimator \( \hat{\Xi}_{k,c} \) by (20) using \( \lambda \) from clustering step above and transform it back to local spot covolatility estimator \( \hat{\Sigma}_{k,\text{glasso}} = D_c(\hat{\Xi}_{k,c})^{-1}D_c \). Compute the (vectorized) weighted covolatility estimator \( \text{vec}\hat{\Sigma}_{k,c} \) from (11).

(v) Combine the \( M \) LMME estimators (from each cluster) into the spot covolatility estimator of the full set of assets and substitute the zero entries representing covariances among different clusters by the respective elements of the PILOT obtaining \( \hat{\Sigma}_{k,\text{glasso}} \).

(vi) Finally, to insure well-conditioned matrix estimator, initialize a full GLASSO estimation with inverted \( \hat{\Sigma}_{k,\text{glasso}} \) as a starting value.

The final estimator is obtained as:

\[
\text{DnC-LMME} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Sigma}_{k,\text{glasso}}. \tag{21}
\]

The last step on each block, uses the composite estimator as a starting value of the full regularization. The idea was proposed in Hsieh et al. (2012) and was called the divide-and-conquer (DnC) method. In contrast to Tan et al. (2015) the DnC Hsieh et al. (2012) is a well-conditional GLASSO estimator.

It is quite easy to substitute the GLASSO regularization by any other method operating on covariance or correlation matrices. The Sparse Estimator of High-Dimensional Correlation (SEC) proposed by Cui et al. (2016) solves similar constrained optimization problem as GLASSO

\[
\hat{R}_\lambda = \arg\min_{R \in \mathbb{R}^m, F > 0} \frac{1}{2} \| R - \hat{R} \|_F^2 + \lambda \| R \|_1, \tag{22}
\]

however it enforces sparsity directly in the correlation matrix, not in its inverse. The \( E \in \mathbb{R}^{(m \times m)} \) is identity matrix and \((\log_{10} \epsilon) \in (-8, -5)\). As one can see, compared to (20) the objective function contains Frobenius norm. The resulting estimator is then obtained in same way as the DnC-LMME by running the full portfolio regularization on the composite estimator:

\[
\text{SEC-LMME} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Sigma}_{k,\text{sec}}. \tag{23}
\]
4. Quantifying information loss for LMME

The grouping and regularization would, as least in theory, lead to loss of efficiency. Therefore we would like to quantify the information distance between the original LMME proposed by Bibinger et al. (2014) and the adjusted LMMEs form previous section. Tumminello et al. (2007, 2010) proposed a method how to measure such distance between estimators of correlations matrices which exploit regularization and patterns recognition techniques (e.g., shrinkage clustering, networks and trees). The measure is based on Kulback-Leibler (K-L) distance between two densities $p$ and $q$:

$$K(p, q) = \mathbb{E}_p \left( \log \frac{p}{q} \right), \quad (24)$$

which under normality and zero-mean & unit-variance assumption boils down to:

$$K(R_1, R_2) = \frac{1}{2} \left( \log \left( \frac{\det R_2}{\det R_1} \right) + \text{tr} \left( R_2^{-1} R_1 \right) - m \right). \quad (25)$$

Note that
- $(25)$ is $\geq 0$ with equality iff $R_1 = R_2$.
- The distance is asymmetric in the densities $p$ and $q$. Therefore, we keep $R_1$ fixed and plug the adjusted estimators in for $R_2$.
- Both $R_1, R_2$ must be positive definite and the measure would become numerically unstable if the arguments were ill-conditioned.
- If $R_2 = \mathbb{E}_m$, $(25)$ reduces to $K(R_1, R_2) = -\frac{1}{2} \log(\det R_1) \geq 0$ since $\det(R) \in [0, 1]$.

Tumminello et al. (2007) showed that if observations are iid $x_1, \ldots, x_n \sim N_m(0, R)$

$$\mathbb{E} \left[ K(\hat{R}, R) \right] = \frac{1}{2} \left( m \log \left( \frac{n}{2} \right) - \sum_{t=n-m+1}^{n} \psi(t/2) \right) \quad (26)$$

where $\hat{R}$ is the maximum likelihood (ML) estimator and $\psi$ is poly-gamma function. The value $(26)$ can be conveniently used as reference distance between the maximum likelihood estimator $\hat{R}$ and the true correlation matrix $R$. Note that $(26)$ is independent of $R$. The intuition on how to use $(26)$ is following. Under the curse of dimensionality, ML estimator might become noisy. Hence a regularized or otherwise adjusted counterpart of MLE might recover true $R$ better. Should it be the case, then plugging $R_1 = \hat{R}$ and $R_2 = \hat{R}^{adj}$ into $(25)$ should give values close to $(26)$. Hence, if $(25)$ gives smaller values than $(26)$ then $\hat{R}^{adj}$ keeps some extra noise. On the contrary, the larger $(25)$ gets compared to $(26)$ the more information has been lost.
Modification for high-frequency observations. Tumminello et al. (2007) counts on usual equidistant Gaussian time series data thus using sample correlation matrix as the MLE. In our case however, the observations are non-equidistant asynchronous realizations of a continuous price process. The frequency-domain approach of Bibinger et al. (2014) however turns these observations into spectral statistics (7) such that:

\[ s_{jk} \sim N_m(0, C_{jk}), \]  

The LMME, in its original version would be a suitable substitution for the ML estimator, however, as we mentioned in the introduction, it is infeasible to obtain LMME from large portfolios. Therefore, we substitute the infeasible LMME with \( \hat{R}_1, \hat{R} = \text{HCC-LMME} \) where the maximal size of a cluster reaches 150 assets. Note that computing LMME for such large clusters is only possible thanks to extra-ordinary computational power provided by the Vienna Scientific Cluster. The role of \( \hat{R}_2 \) in (25) and \( \hat{R} \) in (26) is given to the adjusted (regularized) estimator which is the feasible (regularized) counterpart of HCC-LMME (with maximal cluster size 50) and regularized versions thereof, i.e. DnC-LMME and SEC-LMME.

We present the sample data on SP 500 universe, for which we compute the estimators, in the next section. Here we only summarize the K-L distances, averaged across 993 trading days in our sample, in given in Table 2. As mentioned before, the matrices need to be well-conditioned in order to get stable values of (25). Therefore, we show the results for smoothed matrices over smoothing window of 5 and 20 days. This amount of smoothing provides the matrices with enough stability. We can see that the “closest” to the HCC-LMME is the estimator regularized by divide-and-conquer method for which (25) is approximately twice as large as should be if the estimator would recover true correlation matrix. On the contrary the regularization by SEC obviously filter out much more information from the large-cluster HCC-LMME estimator. The small-cluster non-regularized HCC-LMME is closer to the ML estimator than SEC-LMME but further than divide-and-conquer estimator, hence we can see that regularization based on the precision matrix provides some advantage in terms of information. None of the estimators shows lower values than the benchmark which means that on average, all estimators filter out excessive amounts of information from the data.

5. Portfolio allocation: Setup and empirical results

Our empirical comparison uses mid-quote data on 426 SP 500 constituents. The limit order book quote data were reconstructed by LOBSTER from NASDAQ data feed for the period of 1/1/2012-31/12/2015 (994 trading days). The quotes were further processed according to cleaning procedure by Barndorff-Nielsen et al. (2009) and revised for 0 returns, which significantly reduced the computational burden.
The revisions are used for computation of the PILOT and CHOLCOV estimators described in Section 2 and DnC-LMME and SEC-LMME from Section 3. Additionally, we use the regularization and blocking estimator (RnB) by Hautsch et al. (2012) and 5 minute realized covolatility estimator (RV5min) as established benchmark HF estimators. The RnB is based on 4 liquidity groups as suggested by Hautsch et al. (2015). The curse of dimensionality affects all estimators to some extend, but some require additional regularization. We therefore employ eigenvalue cleaning procedure proposed originally by Laloux et al. (1999) for regularization of PILOT and RnB, whereas RV5min and CHOLCOV turned out well conditioned even without further adjustments. The regularized estimators are further smoothed across 1, 2, 5 and 20 days. The regularization and smoothing has twofold effect on the resulting portfolio weights. First it stabilize the weights, which leads to reduced turnover. The second effect is a classical bias-variance type trade-off. On one hand, it introduces additional bias, but on the other, it can reduce the variance of the estimator and furthermore, smoothing also introduces valuable information from past days. We use open-to-close covariance matrices and returns, hence the overnight effect is treated as deterministic jump. For each day in the sample we compute respective estimator and (after smoothing) we forecast the portfolio weights for the next day. We use the global minimum variance portfolio framework and evaluate the estimators forecasting performance. The optimal portfolio weights (19) solve the optimization problem:

$$
\text{min} \quad \omega_{t,t+1}^T \Sigma_{t,t+1} \omega_{t,t+1}, \quad \text{s.t.} \quad \omega_{t,t+1}^T \iota = 1.
$$

We denote the forecasted weights by $\hat{\omega}_{t,t+1}$ and compute following portfolio characteristics:

- Realized portfolio variance

$$
\sigma_{t,t+1}^2 = \hat{\omega}_{t,t+1}^T \text{MRK}_{t,t+1} \hat{\omega}_{t,t+1},
$$

- Portfolio turnover defined as

$$
\text{po}_{t,t+1} = \sum_{l=1}^m \left| \hat{\omega}_{l,t+1}^T - \hat{\omega}_{l,t-1}^T \frac{1 + r_{l,t-1}}{1 + \text{pr}_{t,t+1}} \right|,
$$

where $\text{pr}_{t,t+1} = \omega_{t-1,t}^T r_{t-1,t}$ is the portfolio return from time $t - 1$ till $t$.

- $L2$-norm of portfolio weights in order to track portfolio concentration which grows if there are many extreme positions (note that minimal value is achieved by equally weighted portfolio.).

$$
\text{pc}_{t,t+1} = \|\omega_{t-1,t}\|_2.
$$
Finally we also report the sum of negative weights

$$p_{s,t+1} = \sum_{l=1}^{m} \omega_{l,t+1} I \{ \omega_{l,t+1} < 0 \},$$  \hspace{1cm} (32)$$

in order to measure the amount of short-selling implied by the respective estimator. Apart from these 4 portfolio characteristics, we compute utility fees for a risk-averse investor with quadratic utility function

$$U(\mathbf{p}_{\text{r},t+1}) = 1 + \mathbf{p}_{\text{r},t+1} - \gamma \frac{\gamma}{2(1+\gamma)} (1 + \mathbf{p}_{\text{r},t+1})^2,$$  \hspace{1cm} (33)$$
defined as $\Delta_\gamma$, such that

$$\sum_{t=1}^{T-1} \mathbb{E} \left[ U(\mathbf{p}_{\text{r},t+1}) \right] = \sum_{t=1}^{T-1} \mathbb{E} \left[ U(\mathbf{p}_{\text{r},t+1} - \Delta_\gamma) \right].$$  \hspace{1cm} (34)$$

The $\Delta_\gamma$ can be interpreted as fees (in basis points) a risk-averse investor is willing to pay in order to switch from estimator I to estimator II. Hautsch et al. (2015) showed that the solution depends on both the expected portfolio return and portfolio variance. For the empirical evaluation, we therefore assume a constant expected portfolio return $\mu = 0.05$. The parameter $\gamma \in \{1, 10\}$ represents the relative risk-aversion of an investor. Hautsch et al. (2015) also incorporated the effect of transaction costs into the economic utility comparison by defining the performance fees net of difference in transaction costs between the two estimators I and II as

$$\Delta^c_\gamma = \Delta_\gamma - c \mathbb{D}_{\text{po}},$$  \hspace{1cm} (35)$$

where $c$ denotes proportional transaction costs on each traded dollar and $\mathbb{D}_{\text{po}} = \overline{\mathbb{p}^\text{II}} - \overline{\mathbb{p}^\text{I}}$. The focus is on so called break-even trading costs, which would imply the $\Delta^c_\gamma = 0$. The interpretation of the quantity is explained in Table 3. The summary of the portfolio characteristics is in Table 4. First we compare the results obtained for our sample period (1/2012-12/2015), with the results of Hautsch et al. (2015) for the pre-crisis period (1/2007-6/2008). Clearly, we can see that the annualized portfolio volatility has decreased by several basis points. For instance the equal weight portfolio for the pre-crisis period (1/2007-6/2008) yield average annual portfolio volatility of $\bar{\sigma}_p^a = 15.24$ bp whereas in our case it merely reaches 10 bp. Similarly for the EC-RnB, Hautsch et al. (2015) reported portfolio volatility between 7 and 10 bp whereas we have values close to 6 bp. The overall decrease in portfolio volatility can be caused by several factors. One of them is the choice of realized covolatility matrix used in (29). Hautsch et al. (2015) used RV5min whereas we use MRK. The difference between the two estimators can be seen in the Figure 1.
For our sample period, the absolute correlations estimated by RV5min are on average 10 times smaller than those by MRK. RV5min requires synchronization of prices among all portfolio constituents. As we pointed out earlier, in our case the differences in liquidity are substantially larger than by Hautsch et al. (2015) who’s portfolio counted 350 most liquid SP 500 constituents obtained from TAQ database. The number of synchronized observations is upper-bounded by the least liquid asset and further reduced by the equidistant (5 min) sampling intervals. Therefore it is possible that there are large gaps between return observations, which result in the low correlations. Another factor that may cause such decrease in portfolio volatility is that we use a relatively stable sample period (at least compared to the second half of the last decade).

On the contrary, compared to the turnover reported by Hautsch et al. (2015), it has increased for the RnB estimator by approx. 50 bp p.a. We assume this is caused by the fact that we work with substantially larger portfolio.

Note that the smoothing of the estimators can help the forecasting performance. It seems that a weekly window yields the best trade-off between additional bias and potentially lower variance of the estimator. Interestingly, CHOLCOV and RV5min do not benefit from smoothing in terms of portfolio volatility. Since smoothing has a regularization effect, it is plausible that these two estimators simply do not benefit from the additional stabilization since they are both already well conditioned even without it. Regrettably, they also do not benefit from the new information from past days. Hence smoothing seems to further increase this discrepancy between strength of correlations estimated by RV5min/CHOLCOV and by the rest of our estimators. On the other hand, the effect of smoothing on portfolio turnover is for all estimators positive, as one would expect.

Comparing the estimators at hand, first thing we notice is relatively poor performance of the CHOLCOV estimator even if compared to RV5min. On the contrary, the PILOT is surprisingly good, especially for the one-week smoothing window. In fact the performance of PILOT is better as by the LMME if the latter is regularized by SEC method. But with DnC, i.e. with focus on the precision matrix, LMME shows some gains over the PILOT both in terms of volatility and turnover. Finally the RnB with 4 liquidity groups performs better than both PILOT and LMME while the latter two overtake RnB as they benefit from additional smoothing over weekly and monthly windows and therefore lead to lower portfolio volatility.

We notice that there is clearly a trade-off between the portfolio turnover and portfolio volatility among the estimators. For instance the inferior performance of CHOLCOV in terms of portfolio volatility can be, up to some extend, compensated by the lowest portfolio turnover among all HF estimators as well as by no (or very little) amount of short selling and almost perfectly balanced portfolios being close to the lower bound.

The economic utility comparison in Table 5i confirms the expected, namely that risk-
averse investor would be willing to switch from equal weighted portfolio to any of the HF based strategies. With the exception of CHOLCOV investors would be willing to pay between 3.9 and 11.5 bp in order to change their strategy from RV5min to the “more sophisticated” estimators which can deal with asynchronicity. Table 5ii further shows that even investor with lower risk aversion would be willing to switch from the PILOT to DnC-LMME and for a monthly smoothing window, the latter holds even under (unrealistic) assumption of transaction benefits. A investor with higher risk aversion would be even willing to pay almost 6.5 bp fees for the change in his strategy. Finally an investor who uses RnB would switch to DnC-LMME only if the estimators that he uses were smoothed across (at least) weekly window. In the latter case, he would be willing to pay up to 4.5 bp fees under relatively high transaction fees per traded dollar.

6. Discussion

We explore the benefits knowing the full correlation structure of a high-dimensional portfolio for high-frequency traders who search for global minimal variance portfolios. Based on the empirical out-of-sample forecasting of optimal portfolio weights, we find evidence that full correlation structure helps reducing the portfolio volatility but leads to higher turnover, more short-selling and possibly unbalanced portfolios, which however a under quadratic utility still results in the willingness of risk-averse investors to pay part of their profit for using efficient estimators.

We had to adjust the local method of moments estimator proposed by Bibinger et al. (2014) in order to accommodate it for high-dimensional portfolios. It has show beneficial to explore regularization on the precision matrix instead of directly on the covariance matrix. The method used in our paper is based on graphical LASSO and leads to improved empirical performance as well as less information loss compared to method which directly reduces the bandwidth of the covariance matrix. Some problems regarding the LMME were not yet properly addressed. In cases such as ours, when estimators are applied locally over intra-day intervals, common situation occurs, that there are less than 3 observations on a interval. In the current form neither PILOT or LMME account of this fact. The modification of the estimators to account for different intensities of trading at the beginning/end and in the middle of the trading day is definitely of practical importance and should be addressed in the future. Additionally, the GLASSO as applied here is disentangled into a clustering and regularization step. A different approach proposed by Pavlenko et al. (2012) applies GLASSO in order to estimate a sparse precision matrix $\Xi$ followed by the Cuthill-McKee algorithm to enforce approximate block-diagonal structure for the precision matrix. This alternative way of obtaining the blocks has it drawbacks due to which we abstain from its use. First the Cuthill-McKee algorithm enforces approximate
block-diagonal structure. However, the final decision about number of blocks (and about their size) must be made which is difficult without previous visual inspection. This is of course possible if one has only few matrices, yet, for us, we would have to inspect possibly hundreds of \((m \times m)\) matrices for each trading day, which is infeasible. Yet, if we find a method to identify the blocks efficiently, the approach could provide a good alternative to the currently described HCC.

References


## Tables and Figures

<table>
<thead>
<tr>
<th>linkage</th>
<th>description</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>equal weight average distance</td>
<td>preferred by Tumminello</td>
</tr>
<tr>
<td>centroid</td>
<td>centroid distance</td>
<td>no try</td>
</tr>
<tr>
<td>complete</td>
<td>furthest distance</td>
<td>similar to average</td>
</tr>
<tr>
<td>median</td>
<td>weighted center of mass distance</td>
<td>no try</td>
</tr>
<tr>
<td>single</td>
<td>shortest distance</td>
<td>bad</td>
</tr>
<tr>
<td>weighted</td>
<td>weighted average distance</td>
<td>no try</td>
</tr>
</tbody>
</table>

Table 1: Linkage functions used for hierarchical clustering.

<table>
<thead>
<tr>
<th>Smoothing</th>
<th>hcclmm</th>
<th>hcclmm-dc</th>
<th>hcclmm-sec</th>
<th>eqw</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>day</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>week</td>
<td>2.42</td>
<td>1.97</td>
<td>4.61</td>
<td>2.86</td>
</tr>
<tr>
<td>month</td>
<td>3.13</td>
<td>2.76</td>
<td>6.07</td>
<td>6.04</td>
</tr>
</tbody>
</table>

Table 2: Kulback-Leibler distance between the HCC-LMM estimator and its adjustments relative to the reference value (26).
Figure 1: Cross-sectional averages of elements from smoothed 5 min realized covolatility (top) and MRK (bottom) estimates using daily, weekly and monthly smoothing windows. On the left are averaged volatilities as annualized square roots of diagonal elements reported in percentage points. On the right are averaged absolute correlations.
Table 3: Interpretation of break even transaction costs $c = \frac{\Delta \gamma}{D_p \gamma}$, i.e. proportional transaction costs on each traded dollar for which the utility (net of difference in transaction costs) of a risk-averse investor from switching the investment strategy is equal 0.

<p>| $\Delta \gamma$ | $D_p \gamma$ | interpretation of $|c|$ |
|-----------------|----------------|---------------------|
| $&gt; 0$           | $&gt; 0$          | maximal transaction costs by which the investor still want to switch |
| $&gt; 0$           | $&lt; 0$          | minimal transaction benefits by which the investor no more wants to switch |
| $&lt; 0$           | $&gt; 0$          | minimal transaction benefits by which the investor would switch |
| $&lt; 0$           | $&lt; 0$          | minimal transaction costs by which the investor would switch |</p>
<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\tilde{\sigma}_p^a$</th>
<th>$\bar{p}_o$</th>
<th>$\tilde{s}_p$</th>
<th>$\bar{p}_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pilot-ec(1)</td>
<td>6.28</td>
<td>240.73</td>
<td>-0.46</td>
<td>0.21</td>
</tr>
<tr>
<td>pilot-ec(2)</td>
<td>5.81</td>
<td>214.30</td>
<td>-0.71</td>
<td>0.24</td>
</tr>
<tr>
<td>pilot-ec(5)</td>
<td>5.49</td>
<td>177.67</td>
<td>-1.12</td>
<td>0.30</td>
</tr>
<tr>
<td>pilot-ec(20)</td>
<td>5.54</td>
<td>115.43</td>
<td>-1.68</td>
<td>0.35</td>
</tr>
<tr>
<td>hccclmm-sec(1)</td>
<td>7.25</td>
<td>151.02</td>
<td>-0.16</td>
<td>0.13</td>
</tr>
<tr>
<td>hccclmm-sec(2)</td>
<td>6.97</td>
<td>128.82</td>
<td>-0.27</td>
<td>0.15</td>
</tr>
<tr>
<td>hccclmm-sec(5)</td>
<td>6.77</td>
<td>111.75</td>
<td>-0.49</td>
<td>0.17</td>
</tr>
<tr>
<td>hccclmm-sec(20)</td>
<td>7.02</td>
<td>77.00</td>
<td>-0.80</td>
<td>0.20</td>
</tr>
<tr>
<td>hccclmm-dc(1)</td>
<td>6.20</td>
<td>249.76</td>
<td>-0.47</td>
<td>0.20</td>
</tr>
<tr>
<td>hccclmm-dc(2)</td>
<td>5.77</td>
<td>225.70</td>
<td>-0.73</td>
<td>0.23</td>
</tr>
<tr>
<td>hccclmm-dc(5)</td>
<td>5.41</td>
<td>180.88</td>
<td>-1.10</td>
<td>0.27</td>
</tr>
<tr>
<td>hccclmm-dc(20)</td>
<td>5.42</td>
<td>101.31</td>
<td>-1.44</td>
<td>0.29</td>
</tr>
<tr>
<td>cholcov-ec(1)</td>
<td>8.33</td>
<td>59.62</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>cholcov-ec(2)</td>
<td>8.38</td>
<td>32.62</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>cholcov-ec(5)</td>
<td>8.46</td>
<td>14.28</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>cholcov-ec(20)</td>
<td>8.70</td>
<td>4.06</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>rb4-ec(1)</td>
<td>6.07</td>
<td>260.02</td>
<td>-0.56</td>
<td>0.22</td>
</tr>
<tr>
<td>rb4-ec(2)</td>
<td>5.70</td>
<td>235.78</td>
<td>-0.84</td>
<td>0.26</td>
</tr>
<tr>
<td>rb4-ec(5)</td>
<td>5.55</td>
<td>194.86</td>
<td>-1.23</td>
<td>0.31</td>
</tr>
<tr>
<td>rb4-ec(20)</td>
<td>6.08</td>
<td>98.36</td>
<td>-1.46</td>
<td>0.32</td>
</tr>
<tr>
<td>rv-5min(1)</td>
<td>6.88</td>
<td>95.33</td>
<td>-0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>rv-5min(2)</td>
<td>6.89</td>
<td>55.62</td>
<td>-0.21</td>
<td>0.12</td>
</tr>
<tr>
<td>rv-5min(5)</td>
<td>6.94</td>
<td>26.06</td>
<td>-0.21</td>
<td>0.12</td>
</tr>
<tr>
<td>rv-5min(20)</td>
<td>7.24</td>
<td>8.29</td>
<td>-0.21</td>
<td>0.11</td>
</tr>
<tr>
<td>eqw</td>
<td>9.89</td>
<td>0.81</td>
<td>0.00</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 4: Forecasted global minimum portfolio characteristics measuring performance of different high-frequency co-volatility estimators. Portfolio constituents are 426 assets which were in SP 500 index over the whole sample period 1/1/2012 - 31/12/2015. The optimal portfolio weights are obtained from each estimator for horizon $h = 1$ day. The portfolio characteristics are annualized portfolio volatility ($\tilde{\sigma}_p^a$) obtained using Multivariate realized kernel estimator as future realized co-volatility measure, portfolio turnover as proxy for transaction costs ($\bar{p}_o$), $L_2$ norm of the portfolio weights as measure of portfolio concentration ($\bar{p}_c$) and sum of negative portfolio weights as measure of the amount of short selling ($\tilde{s}_p$). The reported values are averages across sample period. Annualized volatility $\sigma$ and $p_o$ are given in basis points (100 bp =1%). The estimators are further smoothed over 1, 2, 5 or 20 days. The PILOT, LMME and CHOLCOV are computed first locally on $B \leq 5$ intra-day intervals with number of intervals chosen by proportionality parameter $\theta_h = 0.2$. The spectral cut-off ($J \leq 5$) is chosen by $\theta_J = 8$. For hierarchical correlation clustering, the number of clusters is $\leq 300$ and size of each cluster is $\leq 50$. The RnB estimator is based on 4 liquidity groups. All estimators (except of 5 min realized co-volatility) are regularized before their are smoothed. We also report characteristics based on equal weighted portfolio.
Table 5: The \((\Delta_y^\gamma)\) measure fees a risk-averse investor with quadratic utility and relative risk aversion \(\gamma \in \{1, 10\}\) would pay to switch between respective estimators. Moreover, we report the break-even transaction costs \(c^*\) in basis points, defined as the ratio of the fees \((\Delta_y^\gamma)\) and the difference of average portfolio turnovers. The forecasting horizon \(h = 1\) day. Sample period is 1/1/2012 - 31/12/2015. We assume a constant conditional mean return identical across all stocks \(\mu = 0.05\).