

# **Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity**

**Kris Boudt, Sébastien Laurent, Asger Lunde  
and Rogier Quaadvlieg**

**CREATES Research Paper 2014-5**

# Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity.

This Version: February 24, 2014

Kris Boudt<sup>a,b</sup>, Sébastien Laurent<sup>c</sup>, Asger Lunde<sup>d,e</sup>, Rogier Quaadvlieg<sup>f,\*</sup>

<sup>a</sup>*Department of Business, Vrije Universiteit Brussel, Belgium.*

<sup>b</sup>*VU University Amsterdam, Netherlands.*

<sup>c</sup>*Aix-Marseille Univ. (Aix-Marseille School of Economics), CNRS & EHESS, France*

<sup>d</sup>*CREATES, Aarhus University, Denmark*

<sup>e</sup>*Department of Economics and Business, Aarhus University, Denmark*

<sup>f</sup>*Department of Finance, Maastricht University, Netherlands*

---

## Abstract

An estimator of the ex-post covariation of log-prices under asynchronicity and microstructure noise is proposed. It uses the Cholesky factorization on the correlation matrix in order to exploit the heterogeneity in trading intensity to estimate the different parameters sequentially with as many observations as possible. The estimator is guaranteed positive semidefinite. Monte Carlo simulations confirm good finite sample properties. In the application we forecast portfolio Value-at-Risk and sector risk exposures for a portfolio of 52 stocks. We find that forecasts obtained from dynamic models utilizing the proposed high-frequency estimator provide statistically and economically superior forecasts to models using daily returns.

*Keywords:* Cholesky decomposition, Integrated covariance, Non-synchronous trading, Positive semidefinite, Realized covariance.

JEL: C10, C58

---

## 1. Introduction

The availability of high-frequency data and a large variety of estimators harnessing their information, has led to much greater understanding of the covariation between financial time series. The estimation of covariance matrices is vital in many interesting and important financial, economic and statistical applications. However, most of these do not only require the estimated matrix to be accurate, but also positive semidefinite (PSD). Unfortunately, in search of the former, many proposed

---

<sup>\*</sup>This research was financially supported by a grant of the Dutch Organization for Scientific Research (NWO). CREATES (Center for Research in Econometric Analysis of Time Series) is funded by the Danish National Research Foundation under grant number DNRF78. We would like to thank participants at VU University Seminar, Netherlands Econometric Study Group 2013, the 13th OxMetrics User Conference and the 7th International MIFN workshop. The authors thank Francesco Violante for sharing his code for the RDCC model and Mark Podolskij for useful comments.

<sup>\*</sup>Corresponding author: Department of Finance, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.

*Email addresses:* kris.boudt@econ.kuleuven.be (Kris Boudt), sebastien.laurent@univ-amu.fr (Sébastien Laurent), alunde@econ.au.dk (Asger Lunde), r.quaadvlieg@maastrichtuniversity.nl (Rogier Quaadvlieg)

estimators have sacrificed the latter (e.g. Hayashi and Yoshida, 2005; Zhang, 2011; Lunde et al., 2012).

Estimation of realized covariance at ultra high frequencies is made difficult by two empirical phenomena which induce biases in the estimates. First, the presence of market micro-structure noise (e.g. bid-ask bounce), and second, non-synchronous trading. Up until recently, lower frequency returns, for instance every 20 minutes, were used to avoid these biases. Now, estimators that are robust to both problems have been proposed, and the only limit to the frequency of data one faces is due to the fact the observations have to be synchronized.

The multivariate realized kernel of Barndorff-Nielsen et al. (2011) uses refresh-time sampling to synchronize data. Although this synchronization technique is quite efficient in terms of observations retained, its problem is that the number of observations is always determined by the least frequently traded asset. To diminish that effect, several papers try to make more efficient use of data by splitting up estimation into subsets of the data. Hautsch et al. (2012) propose a method that applies the multivariate kernel to separate groups of liquid and illiquid assets, and combines these estimates into a full matrix. Aït-Sahalia et al. (2010) synchronize pairs and use the so-called polarization result to estimate the covariance bivariate, using univariate estimators. Lunde et al. (2012) use a class of Composite estimators, which estimate the variances univariately, and use bivariate sampling to obtain correlations. Fan et al. (2012) use both the polarization and pairwise refresh-time techniques. These estimators increase efficiency by using more observations, but sacrifice positive semidefiniteness in the process.

The contribution of this paper is a method that estimates each element sequentially, whilst ensuring the final estimate to be positive semidefinite. Because our estimator of the Integrated Covariance is based on the Cholesky decomposition, we denote it CholCov. By applying an orthogonal decomposition to the covariance matrix we reduce the estimation from a  $d$ -dimensional matrix to a large number of bivariate estimations on transformed series, and ensure a well-conditioned matrix. However, due to the necessary transformations, we cannot sample over just two series for each element, but have to iteratively sample over an increasing set of series.

Each bivariate estimate on the transformed series can be done with any currently available or future estimator, as long as it is robust to the biases induced by ultra high-frequency data. Suitable candidate estimators include the sub-sampling estimator (Zhang, 2011; Boudt and Zhang, 2014), pre-averaging estimator (Christensen et al., 2010), kernel estimator (Barndorff-Nielsen et al., 2011) and likelihood based method (Aït-Sahalia et al., 2010), as well as recent advances by Peluso et al. (2012) and Shephard and Xiu (2013). In this paper we use the pre-averaging estimator.

In an extensive simulation study, we find our estimator performs at least as well as its main competitors, and it offers significant improvements when estimating vast-dimensional matrices and/or in scenarios with high heterogeneity in trading frequencies. Its performance is similar to the Composite versions without resorting to an arbitrary rotation to make the estimate positive semidefinite.

For the empirical application, we use our estimator to forecast portfolio Value-at-Risk and to estimate industry exposures for 52 assets. We apply recent advances in the literature on the dynamic modeling Integrated Covariances on the CholCov, and compare the forecasting accuracy to those obtained using dynamic specifications on daily returns only. We find that models utilizing the CholCov in their estimation greatly improve both unconditional coverage and independence between Value-at-

Risk violations, and that they lead to the most informative sector exposure forecasts amongst our models considered.

The paper is structured as follows. In Section 2 we present the theoretical setup and outline the decomposition. Section 3 first discusses practical issues in preparing the data for estimation, and then presents the algorithm that delivers the estimate. In the next section we outline its general asymptotic properties, which depend on the candidate estimator. We provide an example for the pre-averaging estimator. Section 5 presents a summary of some extensive Monte Carlo simulations designed to highlight the improvements due to more efficient data-sampling. Sections 6 and 7 provide the empirical application on the forecasting of portfolio Value-at-Risk and the estimation of industry exposures using the CholCov. Finally, Section 8 concludes. An accompanying webappendix presents an alternative estimator, the PalCov, and additional simulation results.<sup>1</sup>

## 2. Theoretical setup

Our aim is to accurately estimate the Integrated Covariance (ICov) matrix of a  $d$ -dimensional Brownian semimartingale process  $Y = (Y^{(1)}, \dots, Y^{(d)})'$ . The measurement is complicated by the fact that the component processes are observed at irregular and non-synchronous time points, and that the price process of interest is observed with measurement error. The actual observed log-prices are denoted  $X = (X^{(1)}, \dots, X^{(d)})'$ . Let the set of all series be denoted  $\mathcal{D}$ , with subsets  $\mathcal{d} \subseteq \mathcal{D}$ . Each component process can be observed at different time points over the interval  $[0, T]$ . For simplicity we take  $T = 1$  in this paper. The observation times for the  $k$ -th asset are denoted by  $0 \leq t_1^{(k)} \leq t_2^{(k)} \leq \dots \leq t_{N(k)}^{(k)} \leq 1$ .  $X$  is driven by the efficient log-price  $Y$ , a Brownian semimartingale defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t^0), P^0)$ :

$$Y(t) = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s), \quad (1)$$

where  $\mu$  is a  $d \times 1$  predictable locally bounded drift process,  $W$  is a  $d$ -dimensional vector of independent Brownian motions and  $\sigma$  a  $d \times d$  càdlàg process such that  $\Sigma(s) = \sigma(s)\sigma'(s)$  is the spot covariance matrix of  $Y$  at time  $s$ .

At very high frequencies, microstructure noise leads to a departure from the Brownian semimartingale. As a result we do not observe  $Y$ , but instead a process  $X$  defined as:

$$X_t^{(d)} = Y_t^{(d)} + \epsilon_t^{(d)}, \quad (2)$$

where  $\epsilon_t^{(d)}$  is microstructure noise and  $Y_t^{(d)}$  is the  $d$ -th component of  $Y$ . In this paper,  $\epsilon_t = (\epsilon_t^{(1)}, \dots, \epsilon_t^{(d)})'$  is assumed to be an i.i.d. process independent of  $Y$ , satisfying

$$\mathbb{E}(\epsilon_t) = 0, \quad \mathbb{E}(\epsilon_t \epsilon_t') = \Psi,$$

with  $\Psi$  a positive definite  $d \times d$  matrix. Hansen and Lunde (2006) show that these assumptions can be called to question at very high frequencies. They can be relaxed to allow for dependence on  $Y$  and

---

<sup>1</sup>The web-appendix is available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2383869](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2383869).

the asymptotic theory in this paper is still valid (see Christensen et al., 2010). The assumptions made here are specific for the pre-averaging estimator, and others may be considered when using a different candidate estimator.

Our parameter of interest is the integrated covariance over the unit interval:

$$\text{ICov} = \int_0^1 \Sigma(s) ds. \quad (3)$$

Estimation of the off-diagonal elements of the ICov requires synchronization of the data. Two approaches are popular. One is to synchronize all observations jointly, but this has the disadvantage of letting the resulting observation frequency be determined by the least liquid asset (e.g. Barndorff-Nielsen et al., 2011). The second approach is to estimate the extra-diagonal elements of the ICov separately using synchronized pairs of price processes, but the result is not ensured positive semidefinite (e.g. Lunde et al., 2012). We propose a third approach that uses a factorization of the spot correlation matrix, combining the advantages of both methods.

### 2.1. Decomposition

For the estimation of the Integrated Covariance, it will reveal useful to decompose each spot covariance matrix into a symmetric factorization of square matrices. Such a factorization is ensured to be positive semidefinite. Additionally, we wish to separate volatilities from correlations and that the parameters defining the spot correlation between  $k$  and  $l$  do not depend on the parameters defining  $m$  and  $n$  (with  $m > k$  and  $n > l$ ). The latter condition ensures that we can estimate the parameters sequentially, which allows more efficient data usage. This is the case for the Cholesky decomposition, but also for the factorization proposed by Palandri (2009). We will focus on the former, and the latter is discussed in the web Appendix.

The Cholesky decomposition is also used in the volatility context in Chiriac and Voev (2011) and Tsay (2010). Chiriac and Voev (2011) decompose the covariance matrix and use time series models to forecast the elements. This way they ensure positive semidefiniteness of covariance forecasts. Tsay (2010) uses the re-parametrization in a multivariate GARCH framework. The use of the Cholesky decomposition is the only similarity with our work. Their methods and goals are different from ours. We propose an ex-post covariance estimator, not a time-series volatility model.

First, we rewrite the spot covariance matrix by rewriting the spot covariances into the product of spot volatilities and spot correlations as in Boudt et al. (2012). The spot correlation between log-returns of assets  $k$  and  $l$  is defined as

$$\rho_{k,l}(s) = \frac{\Sigma_{k,l}(s)}{\sqrt{\Sigma_{k,k}(s)\Sigma_{l,l}(s)}}, \quad (4)$$

where  $\Sigma_{k,l}(s)$  is element  $(k, l)$  of the spot covariance matrix  $\Sigma(s)$ , implying

$$\Sigma_{k,l}(s) = \sigma_k(s)\rho_{k,l}(s)\sigma_l(s)$$

with  $\sigma_k(s) = \sqrt{\Sigma_{k,k}(s)}$ . In multivariate notation, this leads to  $\Sigma(s) = D(s)R(s)D(s)$ , where  $D(s)$  is

a  $d \times d$  diagonal matrix containing the spot volatilities and  $R(s)$  the  $d \times d$  spot correlation matrix.

Second, the spot correlation matrix is split up using the well known Cholesky decomposition, i.e.

$$R(s) = H(s)G(s)H(s)', \quad (5)$$

where  $H(s)$  is a lower diagonal matrix with ones on the diagonal, and  $G(s)$  a diagonal matrix. More specifically,

$$H(s) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h_{21}(s) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{d1}(s) & h_{d2}(s) & \cdots & 1 \end{bmatrix} \quad G(s) = \begin{bmatrix} g_{11}(s) & 0 & \cdots & 0 \\ 0 & g_{22}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{dd}(s) \end{bmatrix}.$$

For instance, omitting the time-dependence, for  $d = 3$ :

$$R = HGH' = \begin{bmatrix} g_{11} & h_{21}g_{11} & h_{31}g_{11} \\ h_{21}g_{11} & h_{21}^2g_{11} + g_{22} & h_{21}h_{31}g_{11} + h_{32}g_{22} \\ h_{31}g_{11} & h_{21}h_{31}g_{11} + h_{32}g_{22} & h_{31}^2g_{11} + h_{32}^2g_{22} + g_{33} \end{bmatrix}. \quad (6)$$

Hence, the elements in the  $H$  and  $G$  matrices are linked to the elements of  $R$  as follows

$$g_{kk} = R_{kk} - \sum_{m=1}^{k-1} h_{k,m}^2 g_{mm} \quad (7)$$

$$h_{kl} = \frac{1}{g_{kk}} \left( R_{kl} - \sum_{m=1}^{k-1} h_{km} h_{lm} g_{mm} \right), \quad (8)$$

for  $k > 1$  with  $g_{11} = R_{11}$ . The  $(k, l)$ -th element of the correlation matrix only depends on  $g_{mm}$  and  $h_{np}$  with  $m, n, p \leq \max(k, l)$ . The elements can therefore be estimated sequentially.

### 3. Estimation

In order to apply the Cholesky decomposition in estimating the ICov, we first need to define how the returns are computed and, since the procedure will be sequential, how the covariance matrix is ordered. Subsection 3.1 reviews the concept of refresh time returns and Subsection 3.2 introduces the liquidity criterion that we use to structure the covariance matrix. Based on these ingredients, Subsection 3.3 defines the CholCov estimator.

#### 3.1. Dealing with asynchronicity

In high-frequency data, assets are traded at irregular intervals and seldom simultaneously. Moreover, not all stocks are equally liquid, such that the number of fresh prices within a fixed time interval may substantially differ across firms. One way to synchronize the data is by means of refresh-time sampling, as proposed by Harris et al. (1995). It picks refresh-times at which all assets have traded at

least once since the last refresh-time point. We use the notation of Hautsch et al. (2012) and define refresh-time for  $t \in [0, 1]$  as follows. The first refresh-time occurs at  $\tau_1 = \max(t_1^{(1)}, \dots, t_1^{(d)})$ , and subsequent refresh-times at  $\tau_{j+1} = \operatorname{argmin}(t_k^{(k)} | t_k^{(k)} > \tau_j, \forall k)$ . The grid, denoted  $\tau^d$ , is a function of the series over which they are sampled, with  $d \subseteq \mathcal{D}$ . The grid's number of observations is denoted  $N^d$ , while the individual series have size  $N^{(k)}$ . Finally, denote the durations as  $\Delta_j^d = \tau_j^d - \tau_{j-1}^d$ . Throughout the paper, for clarity of notation, the superscript may be omitted if it is clear which grid is being discussed, or in general statements concerning every grid.

As a result of the sampling scheme,  $N^d \leq \min_k N^{(k)}$ , and may be a lot smaller depending on the trading pattern and number of series under consideration. Hautsch et al. (2012) illustrate that the percentage data loss can exceed 90% when the number of assets becomes large and arrival rates are unequal. The sample size is clearly largely determined by the least liquid stock. Including a single non-liquid stocks may therefore drastically reduce the estimation efficiency of all elements, including those for which a lot of data is available.

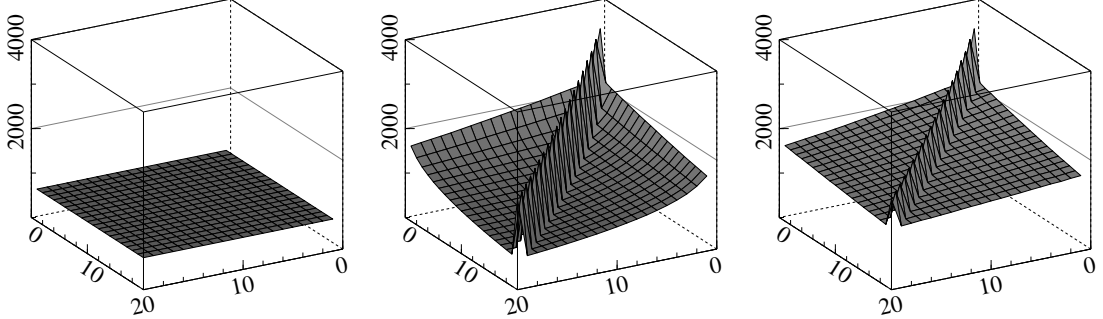
This problem is circumvented by the composite estimation technique used in Lunde et al. (2012) and Fan et al. (2012). The data loss is reduced, but at the expense of positive semidefiniteness, one of the defining properties of a covariance matrix. Moreover, many applications, such as principal components analysis and portfolio optimization, critically rely on it. Therefore the question arises how to project the symmetric matrix onto the space of PSD matrices.

There exist many ways to transform the matrix to a PSD alternative. A general method called shrinkage is proposed in Ledoit and Wolf (2003). In the Realized Covariance setting Barndorff-Nielsen and Shephard (2004b) and Fan et al. (2012) set all negative eigenvalues in the spectral decomposition to zero. Hautsch et al. (2012) impose more structure by employing eigenvalue cleaning, a random matrix theory technique similar in intuition to shrinkage. While such eigenvalue cleaning may increase the efficiency of the covariance matrix estimate, it still remains that for portfolio optimization purposes the dependence of the optimized portfolio on the eigenvalue cleaning is highly undesirable. Schmelzer and Hauser (2013) discuss the severe implications of a negative eigenvalue in the context of portfolio optimization and point to the above mentioned strategies to get rid of them. They conclude, however that, from their experience, a careful analysis of the estimation process itself adds far more value to the trading strategy.

Here, we use the Cholesky decomposition to exploit the heterogeneity in trading intensities and estimate the different parameters sequentially, using as much data as possible. The decomposition holds for any PSD matrix, which is by definition true for the correlation matrix, and the recombined results will also be PSD. However, ensuring positive semidefiniteness does come at a cost. We cannot synchronize just pairs of data, but have to sample over a growing set of series. The first elements are estimated by sampling over two series, but the last elements to be estimated require a common grid on all series. To make optimal use of the data, it is therefore crucial to order the series in terms of decreasing liquidity.

To illustrate, Figure 1 plots a simulated example of the number of observations used for each element using the three different methods. We consider 20 assets, where each asset is slightly less liquid than the last. The leftmost picture depicts a situation in which refresh-time sampling is applied to all assets at once. The rightmost picture depicts pairwise sampling, the situation for the Composite

Figure 1: Number of observations available for each element in estimation of  $\Sigma$ , sampling over all series (left), sequential sampling (middle) and pairwise sampling (right).



*Note:* These graphs depict simulation based number of observations available for the estimation of elements in a  $20 \times 20$  covariance matrix for different sampling schemes. The sampling frequencies follow independent poisson processes with parameter  $\lambda$  ranging uniformly from 10 and 15. For more details, see Section 5. From left to right: Sampling over all series, sequential sampling and pairwise sampling.

estimators of Lunde et al. (2012). The diagonal elements are estimated sampling over just the single series, and therefore also depict the number of observations available for that asset. The off-diagonal elements are estimated using bivariate sampling. The data-loss is therefore minimal. The middle graph depicts sequential sampling, the technique used for our estimator. The diagonal elements are estimated in the same fashion as for the Composite estimator, but we require sampling over more series for the off-diagonal estimates. As such, the number of observations for all elements involving the first asset coincide with those for pairwise sampling, while the observation count for the other elements lies somewhere in between the two other sampling approaches.

### 3.2. Ranking stocks based on liquidity

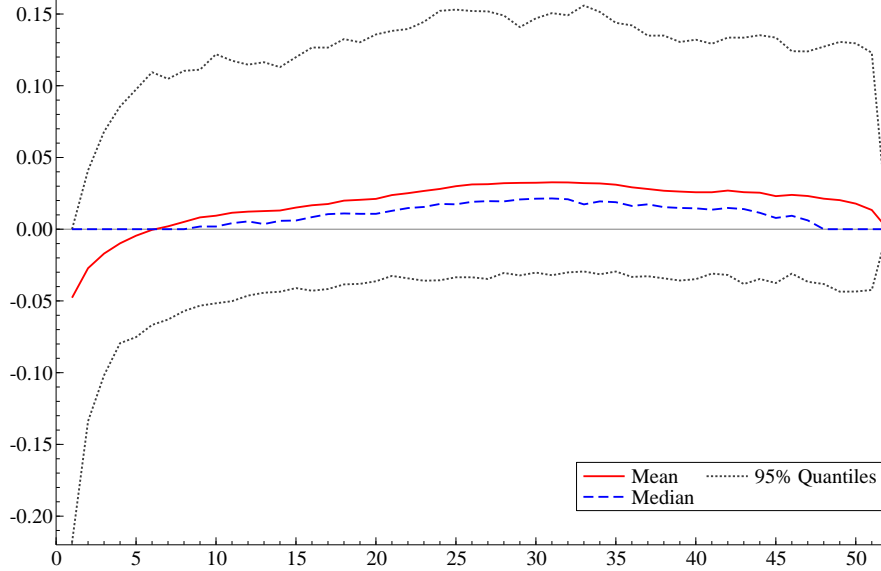
In order to best utilize the sequential estimation, we propose to sort the series in terms of liquidity. For the Regularization and Blocking ('RnB') estimator, Hautsch et al. (2012) sort series in terms of liquidity solely based on the number of observations  $N^{(k)}$ . However, there are many scenarios where this would not lead to a large number of observations when sampling over many series. Instead, we propose the following new liquidity criterion:

$$0 \leq \sum_{j=0}^{N^{(k)}} \left( \Delta_j^{(k)} - \frac{1}{\max_d N^{(d)}} \right)^2 < 1, \quad (9)$$

which is not just based on the number of intraday returns available, but also the degree to which the returns are spread out over the day. To achieve this we compare the time between trades, denoted  $\Delta_j$ , with an optimal grid, defined as the equidistant grid based on the maximum number of observations available for any of the series. The series are ordered in increasing value of the liquidity criterion, such that the first series is the most liquid. We call our method liquidity sorting, as opposed to observation count sorting proposed in Hautsch et al. (2012).



Figure 2: Distribution of log-differences in number of observations between liquidity and observation count sorting, as a function of the number of series synchronized.



*Note:* The graph shows the percentage gains in number of observations from using liquidity sorting instead of observation count sorting on refresh-time samples over an increasing number of series from 1 to 52.

To illustrate the effect of liquidity sorting compared to observation count sorting, we preview results of our empirical application where we apply sequential refresh-time sampling on 52 assets. Figure 2 shows the increase in number of observations based on liquidity sorting compared to sorting purely on number of individual observations. The x-axis denotes the number of series that are synchronized. The graph shows the distribution of the log-difference of the number of observations between the two sampling techniques. As such the y-axis can be interpreted as percentage increases in number of observations of liquidity sorting compared to observation count sorting.

It is clear that the liquidity criterion (9) offers a smoothing of observation count across the increased set of series. Some observations are lost in the beginning of the sequential estimation, but this is at a point where plenty of observations are available, and efficiency is not so much an issue. However, starting from about seven series, the synchronized observation count for liquidity sorting is on average higher by 2 to 3%. Additionally, the distribution is skewed towards higher gains than losses, with the liquidity sorting offering observation increases of 10 to 15% in the 2.5% best cases, compared to losses of around 5% in the worst 2.5% cases. Obviously, when all the series are synchronized, there is no difference.

### 3.3. The CholCov

This section presents our estimator of the Integrated Covariance, which achieves positive semidefiniteness by means of the Cholesky decomposition. Recall that we assume log-prices to follow (1). For ease of exposition, we first consider a simple scenario in which we observe equispaced returns on a common grid, e.g. five minute returns, uncontaminated by microstructure noise, and with constant volatility and correlation throughout the day. We then discuss the necessary adjustments to deal with asynchronicity and microstructure noise, and finally allow for intraday patterns in volatility and correlations.

*No Asynchronicity and no Noise.* First, as volatilities and correlations are assumed constant throughout the day and we focus on one day, we will write  $D(s)$  as  $D$  in this subsection to simplify notation, and use the same notation for matrices  $R$ ,  $G$ , and  $H$ .

Recall we decompose the covariance matrix into volatilities and correlations, which are estimated separately. The first step is to estimate the variances, which are then used to standardize the return series on which we estimate the correlation matrix. Define the series of equispaced intra-day returns  $r_{t_j}^{(1)}, \dots, r_{t_j}^{(d)}$ , where  $r_{t_j}^{(k)} = X_{t_j}^{(k)} - X_{t_{j-1}}^{(k)}$ . Under this scenario, the Integrated Variance (IV) of the series can be estimated consistently by the Realized Variance (RV). The estimates lead to the diagonal matrix  $\hat{D}$ , whose  $(k, k)$ -th element  $\hat{\sigma}_k$  is  $\sqrt{RV}$  of asset  $k$ , with  $N$ , the number of intraday observations. Define  $RV(f^{(k)}) = \sum_{j=0}^N \left(f_{t_j}^{(k)}\right)^2$ , the function that computes the RV based on the series  $f^{(k)}$ .

We standardize the returns to have unit Integrated Variance over the day. To achieve this, the returns are pre-multiplied by  $\hat{D}^{-1}$ , which amounts to dividing all returns by the square root of RV. The standardized log-returns are therefore

$$\hat{u}_{t_j}^{(k)} = \frac{r_{t_j}^{(k)}}{\hat{\sigma}_k}. \quad (10)$$

This ensures that the Integrated Variance of the series  $\hat{u}_{t_j}^{(k)}$  equals one in expectation. Even though we are now technically dealing with the estimation of a correlation matrix, this is unlikely to be exactly true in practice. We therefore do not impose this assumption during the estimation, but will standardize afterwards.

Next, the  $g_{kk}$  and  $h_{kl}$  terms in the decomposition of the correlation matrix in (5) are estimated sequentially. Each element requires only a subset of the data series. The sequential estimation follows directly from the triangularity of  $H$ . To see this, omitting time subscripts for clarity, consider the vector of standardized returns  $u$ , which, under the previously stated assumptions and provided returns are sampled sufficiently frequently such that the drift can be ignored, is normally distributed around zero with covariance matrix  $R = HGH'$ . We can further define  $f = H^{-1}u \sim N(0, G)$ . Since  $H$  is triangular, we obtain that each component in the vector of standardized returns  $u = Hf$ , is an explicit function of the components with a lower index. More precisely,  $f^{(1)} = u^{(1)} \sim N(0, g_{11})$ , and for  $k = 2, \dots, d$ ,

$$\begin{aligned} u^{(k)} &= h_{k1}f^{(1)} + h_{k2}f^{(2)} + \dots + h_{k(k-1)}f^{(k-1)} + f^{(k)} \\ f^{(k)} &\sim N(0, g_{kk}). \end{aligned} \quad (11)$$

Based on the sequence in (11), we can thus estimate the elements of  $H$  by the following algorithm of

Ordinary Least Squares (OLS) regressions:

Step 1:

$$\begin{aligned}\text{Set: } f_{t_j}^{(1)} &= \hat{u}_{t_j}^{(1)} \\ \text{Set: } \hat{g}_{11} &= RV(f^{(1)})\end{aligned}$$

Step 2:

$$\begin{aligned}\text{Regress: } \hat{u}_{t_j}^{(2)} &= h_{21}f_{t_j}^{(1)} + \eta_{t_j}^{(2)} \\ \text{Set: } f_{t_j}^{(2)} &= \hat{u}_{t_j}^{(2)} - \hat{h}_{21}f_{t_j}^{(1)} \\ \text{Set: } \hat{g}_{22} &= RV(f^{(2)})\end{aligned}$$

Step 3:

$$\begin{aligned}\text{Regress: } \hat{u}_{t_j}^{(3)} &= h_{31}f_{t_j}^{(1)} + h_{32}f_{t_j}^{(2)} + \eta_{t_j}^{(3)} \\ \text{Set: } f_{t_j}^{(3)} &= \hat{u}_{t_j}^{(3)} - \hat{h}_{31}f_{t_j}^{(1)} - \hat{h}_{32}f_{t_j}^{(2)} \\ \text{Set: } \hat{g}_{33} &= RV(f^{(3)})\end{aligned} \tag{12}$$

$\vdots$

Step  $d$ :

$$\begin{aligned}\text{Regress: } \hat{u}_{t_j}^{(d)} &= \sum_{l=1}^{d-1} h_{dl}f_{t_j}^{(l)} + \eta_{t_j}^{(d)} \\ \text{Set: } f_{t_j}^{(d)} &= \hat{u}_{t_j}^{(d)} - \sum_{l=1}^{d-1} \hat{h}_{dl}f_{t_j}^{(l)} \\ \text{Set: } \hat{g}_{dd} &= RV(f^{(d)}).\end{aligned}$$

Step 1 is merely an initialization of the sequential estimation. The aim of all other steps is to determine an orthogonal projection of series  $u^{(k)}$  on all previously orthogonalized series  $f^{(l)}$ ,  $l = 1, \dots, k-1$ . To achieve this, as per the first equation in each step, we first regress  $u^{(k)}$  on all previous orthogonalized series  $f^{(l)}$ . The parameter estimates  $\hat{h}_{kl}$  are used in the second equation to compute the next orthogonal projection of series  $u^{(k)}$  on all previous factors, denoted by  $f^{(k)}$ . The process ensures that  $f_{t_j}^{(k)} \perp f_{t_j}^{(l)}$  for  $k \neq l$ . By going through the steps, we obtain the elements of  $H$  and  $G$  as  $h_{kl} = h_{kl}$  and  $g_{kk} = RV(f^{(k)})$ .

In this equispaced, no noise scenario, Realized Variance is a consistent estimator of the Integrated (Co)Variance and OLS on intraday returns leads to consistent estimates of the  $h$  elements. In this case, the estimation of the CholCov is relatively straightforward and, as the Cholesky decomposition is exact, identical to the standard Realized Covariance estimator (RCOV) studied in, amongst others, Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004a). The estimation directly follows the steps previously outlined. In step 1,  $g_{11}$  is obtained as the Realized Variance of  $f^{(1)}$ . In step 2, we first obtain  $h_{21}$  as the OLS-coefficient from regressing  $\hat{u}^{(2)}$  on  $f^{(1)}$ . Then we apply the orthogonal transformation, and estimate  $g_{22}$  as the RV of the projected series  $f^{(2)}$ . In step 3, the next two

elements,  $h_{31}$  and  $h_{32}$  are OLS-coefficients from regressing  $\hat{u}^{(3)}$  on  $f^{(1)}$  and  $f^{(2)}$ . The RV of the orthogonal transformation is the  $g_{33}$  element. We repeat this step  $d$  times, until all elements of  $G$  and  $H$  have been estimated.

The matrix  $\hat{Q} = \hat{H}\hat{G}\hat{H}'$  is guaranteed to be a symmetric positive semidefinite matrix, but to be qualified as a correlation matrix, it needs to have ones on the diagonal. We considered two ways to ensure this. First, as in Equation (2.5) of Pelletier (2006) we could impose the constraint

$$g_{ii} = 1 - \sum_{k=1}^{i-1} h_{(i+1)k}^2 g_{kk},$$

such that the diagonal elements in (6) equal 1 by construction. In Pelletier (2006) this is part of an iterative maximum likelihood estimation of the correlation matrix, whereby the joint optimization of all elements ensures the accuracy of the estimates. However, in our case, there is no iteration and simulations showed that the efficiency of resulting estimate is not satisfactory. This is because the integrated variance of the standardized returns equals one just in expectation, and is unlikely to exactly equal one in practise. The second approach, which we use in this paper, is to simply standardize the matrix to unit diagonal, as in Engle (2002),

$$\hat{R} = \text{diag}(\hat{Q})^{-1/2} \hat{Q} \text{diag}(\hat{Q})^{-1/2}.$$

This ultimately leads to the CholCov estimator for the daily integrated covariance matrix:

$$\text{CholCov} = \hat{D}\hat{R}\hat{D}. \quad (13)$$

*With Asynchronicity and Noise.* In practice, we observe tick-data which has to be synchronized and has microstructure noise. To estimate under this scenario some adaptations have to be made, but the general principle is the same. The problem of noise is easily solved by means of a robust estimator. Asynchronicity could be solved by means of a common refresh-time grid on all series, but we exploit the sequential nature of the regressions by sampling over an increasing number of series (sorted with respect to the liquidity criterion (9)). This offers large benefits in terms of data usage, but it also means that different grids are used for every single element.

Having multiple grids means that the standardized returns have to be recomputed on every single grid, taking into account the time between successive observations. Consider a synchronization that led to a time-grid  $\tau$ , with  $N$  observations. Define the actual observations times as  $\tau_{j+1}^{(k)} = \arg\max(t_k^{(k)} | t_k^{(k)} \leq \tau_j)$ . Each return  $r_{\tau_j}^{(k)} = X_{\tau_j}^{(k)} - X_{\tau_{j-1}}^{(k)}$  has variance  $(\tau_j - \tau_{j-1})\sigma_k^2$ . To obtain unit IV on the daily basis, we estimate the IV using some robust estimator,  $\widehat{IV}$ , and standardize the returns using  $\hat{\sigma}_{\tau_j}^{(k)} = \widehat{IV}^{(k)}$ . The standardized log-returns for the grid  $\tau$ , are therefore

$$\hat{u}_{\tau_j}^{(k)} = \sqrt{\frac{1/N}{(\tau_j^{(k)} - \tau_{j-1}^{(k)})}} \frac{r_{\tau_j}^{(k)}}{\hat{\sigma}_{\tau_j}^{(k)}}. \quad (14)$$

Note that for equispaced returns,  $(\tau_j^{(k)} - \tau_{j-1}^{(k)}) = 1/N$ , the standardized return in (14) thus coincide with those in (10). These series have to be recomputed on every grid we consider, as the different

grids will have different refresh-times.

Now, in order to obtain consistent estimates in the presence of asynchronicity and noise, we estimate the quantities,  $g_{kk}$  and  $h_{kl}$ , using robust realized variances and realized betas<sup>2</sup>, using estimators such as the multivariate Kernel (Barndorff-Nielsen et al., 2011) and MRC (Christensen et al., 2010). Using realized beta instead of OLS has the advantage that the minimum number of observations needed for estimation is a lot lower. For OLS, the number of observations of fully synchronized data needs to be greater than the dimension of the problem, a clear restriction for vast matrices. Realized beta does not have this restriction. Additionally, it allows for optimal bandwidth selection of each individual element.<sup>3</sup>

A drawback of realized beta is that it does not take into account dependence between regressors, and hence the estimator is only consistent if the orthogonal transformations  $f^{(\cdot)}$  are indeed orthogonal. If all betas are estimated on the same grid this condition holds by construction. However, as the grid changes for every element, all the  $f^{(\cdot)}$  have to be recomputed on each new grid using preceding estimates of beta. The betas orthogonalize the series on the grid they are estimated on, but are unlikely to perfectly orthogonalize the same series on a different observation grid. To counter this problem and to ensure consistency of the estimator, we achieve perfect orthogonality by re-estimating preceding betas at every new grid. The realized beta obtained using the smallest set of series, the first time it is estimated, is used as the final estimate in  $H$ . As it uses the smallest set, it will have the most observations and is therefore the most efficient. The following algorithm summarizes the procedure, and makes the link to the set of equations (12).

***CholCov Estimation Algorithm.***

1. First sort the series in terms of decreasing liquidity according to (9), such that series 1 is the most liquid, and series  $d$  the least.
2. *Step 1:*  
 Apply refresh-time on  $a = \{1\}$  to obtain the grid  $\tau^a$ .  
 Estimate  $\hat{g}_{11}$  using an IV estimator on  $f_{\tau_j^a}^{(1)} = \hat{u}_{\tau_j^a}^{(1)}$ .
3. *Step 2:*  
 Apply refresh-time on  $b = \{1, 2\}$  to obtain the grid  $\tau^b$ .  
 Estimate  $\hat{h}_{21}^b$  as the realized beta between  $f_{\tau_j^b}^{(1)}$  and  $\hat{u}_{\tau_j^b}^{(2)}$ . Set  $\hat{h}_{21} = \hat{h}_{21}^b$ .  
 Estimate  $\hat{g}_{22}$  using an IV estimator on  $f_{\tau_j^b}^{(2)} = \hat{u}_{\tau_j^b}^{(2)} - \hat{h}_{21} f_{\tau_j^b}^{(1)}$ .
4. *Step 3:*  
 Apply refresh-time on  $c = \{1, 3\}$  to obtain the grid  $\tau^c$ .  
 Estimate  $\hat{h}_{31}^c$  as the realized beta between  $f_{\tau_j^c}^{(1)}$  and  $\hat{u}_{\tau_j^c}^{(3)}$ . Set  $\hat{h}_{31} = \hat{h}_{31}^c$ .  
 Apply refresh-time on  $d = \{1, 2, 3\}$  to obtain the grid  $\tau^d$ .  
 Re-estimate  $\hat{h}_{21}^d$  at the new grid, such that the projections  $f_{\tau_j^d}^{(1)}$  and  $f_{\tau_j^d}^{(2)}$  are orthogonal.

---

<sup>2</sup>The realized beta is computed by estimating a  $2 \times 2$  realized covariance matrix  $\hat{\Sigma}$ . The realized beta is then computed as  $\hat{h}_{kl} = \hat{\Sigma}_{kl} / \hat{\Sigma}_{ll}$  (Andersen et al., 2006).

<sup>3</sup>Note however that in the no asynchronicity, no noise case discussed previously, OLS and Realized Beta are equivalent.

Estimate  $\hat{h}_{32}^d$  as the realized beta between  $f_{\tau_j^d}^{(2)}$  and  $\hat{u}_{\tau_j^d}^{(3)}$ . Set  $\hat{h}_{32} = \hat{h}_{32}^d$ .

Estimate  $\hat{g}_{33}$  using an IV estimator on  $f_{\tau_j^d}^{(3)} = \hat{u}_{\tau_j^d}^{(3)} - \hat{h}_{32}f_{\tau_j^d}^{(2)} - \hat{h}_{31}f_{\tau_j^d}^{(1)}$ .

5. *Step 4 to d:*

Continue in the same fashion by sampling over  $1, \dots, k, l$  to estimate  $h_{lk}$  using the smallest possible set. Re-estimate the  $h_{nm}$  with  $m < n \leq k$  at every new grid to obtain orthogonal projections. Estimate the  $g_{kk}$  as the IV of projections based on the final estimates,  $\hat{h}$ .

This way we again obtain the estimates of  $\hat{D}$ ,  $\hat{G}$ ,  $\hat{H}$ , and as such  $\hat{R}$ , and one can simply compute the CholCov as in (13).

### 3.4. Decomposition of the ICov assuming locally constant correlations

In this section we outline how one could allow for time-varying spot volatility and spot correlations. To do so, we will assume locally constant correlations, whereas the volatility is allowed to be fully time-varying. However, in the remainder of the paper, we use the implementation as discussed in the previous section.

Consider the finest equispaced time grid  $t_i$ , with  $\delta = t_{i+1} - t_i$  and  $i = 1, \dots, N_t$ . The volatility is allowed to be time-varying throughout the day. The correlation is assumed constant throughout sub-periods of the day, but may change between those periods. Without loss of generality we assume these periods to be of equal length,  $\kappa\delta$ . Then, ICov can be approximated as follows:

$$\text{ICov} \approx \sum_{j=1}^{N_t/\kappa} \left[ \int_{t_{\kappa(j-1)}}^{t_{\kappa j}} D(s) R(t_{\kappa(j-1)}) D(s) ds \right], \quad (15)$$

where  $D(s)$  is the diagonal matrix whose  $k$ th element is the spot volatility of asset  $k$  and  $R(t_{\kappa(j-1)})$  is the constant correlation matrix between  $t_{\kappa(j-1)}$  and  $t_{\kappa j}$ . Equation (15) is a discretized version of (3) and converges to ICov under some smoothness conditions when  $N_t \rightarrow \infty, \kappa \rightarrow 0$ , but with  $N_t/\kappa \rightarrow \infty$ .

In practise we are dealing with a refresh-time grid. We assume the spot correlation to be constant over a fixed number,  $\kappa$ , of refresh-times.  $\tau^{\mathcal{D}}$  is the refresh-time grid using all series under consideration, with  $N^{\mathcal{D}}$  observations.  $\kappa$  is defined on this grid such that the spot correlation is constant for the  $\kappa$  observations at times  $\tau_{\kappa(j-1)}^{\mathcal{D}}, \dots, \tau_{\kappa j}^{\mathcal{D}}$  for  $j = 1, \dots, N^{\mathcal{D}}/\kappa$ . In the remainder of this section the sample is partitioned into  $N^{\mathcal{D}}/\kappa$  sub-samples based on transaction times  $\tau_{\kappa j}^{\mathcal{D}}$ , regardless of what transaction grid the sample has. The number of observation within a subsample will generally be larger than  $\kappa$ . When  $N^{\mathcal{D}} \rightarrow \infty$ , the length of the constant correlation windows converges to zero and does not depend on the inter-arrival times of observations.

The time-varying spot volatilities,  $D(s)$  can be estimated using any integrated variance estimator on the returns. Kristensen (2010) obtains the spot volatilities by estimating the Integrated Variance in local windows using kernels. An alternative would be to explicitly estimate the intraday periodicity (see e.g. Boudt et al., 2011; Bos et al., 2012).

The correlations are estimated using the same algorithm as before, but on just the observations within the window  $\tau_{\kappa(j-1)}$  and  $\tau_{\kappa j}$ . This ultimately leads to the CholCov estimator for the daily

integrated covariance matrix:

$$CholCov = \sum_{j=1}^{N^D/\kappa} \left[ \sum_{\tau_{\kappa(j-1)}^D \leq s \leq \tau_{\kappa j}^D} \left[ \hat{D}(s) \hat{R}(\tau_{\kappa(j-1)}^D) \hat{D}(s) \right] (\tau_{\kappa j}^D - \tau_{\kappa(j-1)}^D) \right]. \quad (16)$$

Finally, the question remains how to determine  $\kappa$  and  $\gamma$  practice. Determining their values is a typical bias, variance trade-off. Higher values will reduce variance but increase bias if the correlation or variance is not constant throughout the window. The value should be chosen as high as possible to use as many observations for any one estimate as possible. One way to determine the values is to take the approach adopted in Boudt et al. (2012) by testing for equal correlation or variance between windows. Previous research suggests that patterns in the correlations are highly persistent and intraday patterns are often negligible (Tang, 1995). Volatility however shows clear intraday patterns (Andersen et al., 2007), suggesting the window length for volatilities should be a lot shorter than for correlations, or that returns should be pre-filtered by a periodicity filter.

#### 4. Properties

The CholCov possesses a combination of the most attractive properties of realized covariance estimators. First, it is positive semidefinite and invertible by construction. Each element in the sum of (16) is PSD by virtue of the Cholesky decomposition, and the sum of PSD matrices is PSD. Invertibility is shown by e.g. Miller (1981).

Second, the CholCov offers great flexibility. We can use any estimator to compute the realized variances and betas. It adopts the robustness properties of whatever estimator one uses. For instance, we can use the Modulated Realized Variance and Covariance to compute  $\hat{g}_{kk}$  and  $\hat{h}_{kl}$  respectively, and obtain robustness against noise, asynchronicity and serial dependence in noise. In principle, one could use any combination of estimators, even within one estimation. One could use pre-averaging techniques to estimate the variances, and kernel-based methods to estimate the realized betas.

Next to robustness, the CholCov also adopts asymptotic properties related to the estimators used. First, under some smoothness conditions on the spot covariance, consistency follows directly from consistency of the estimators used for the  $g_{kk}$  and  $h_{lk}$  elements. As long as we only use estimators that are consistent, all estimates are individually consistent, and hence the combined estimator is consistent. To ensure convergence, it is sufficient to require the spot covariance process to be (Riemann) integrable.

##### 4.1. Asymptotic Distribution

We discuss the asymptotic distribution of the estimates in  $\hat{G}$  and  $\hat{H}$ . The consistency is not affected by the sequential estimation, but the asymptotic variance (avar) may be larger. We only consider the distribution of the estimates conditional on no estimation error in previous estimates. In practice, a more accurate estimate may be obtained by bootstrap methods (see e.g. Hounyo et al., 2013). However, as the estimators used are consistent, there is no systematic bias. We conjecture that asymptotic theory for noise-robust estimators may therefore still hold as the impact of previous stage estimation error can be modeled as noise. Lastly, we only consider the distribution of the estimates

in one window of constant correlation. As there is no overlap between blocks, and the blocks are independent, the avar for the entire day is the period length-weighted sum of the avar matrices.

All elements are estimated on transformed series. The transformation obviously affects the stochastic process, but it remains within the class of Brownian semimartingales. The asymptotic variance of every possible estimator depends on the volatility matrix process, so we need to keep track of the volatility process through the transformations. All transformed series are linear combinations of original series. Hence, the volatility process is always of the form

$$D^{-1}(s)A\sigma(s)dW(s) = \sigma^*(s)dW(s),$$

with  $\sigma^*(s) = D^{-1}(s)A\sigma(s)$ , and therefore  $\Sigma^* = \sigma^*(s)(\sigma^*(s))'$ . Matrix  $A$  contains the weights for the linear transformation. For instance, when  $d = 3$ , and we are interested in the avar of  $h_{32}$  we need the volatility process of the series  $r^{(3)}$  and  $s^{(2)}$ , such that

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -h_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The asymptotic theory of the  $g$  and  $h$  elements follows directly from the theory of whichever estimator we use. In Appendix C we report for completeness these results when the pre-averaging estimator, as detailed in Appendix B, is used. Ideally, one would like the distribution of the final correlation estimates, which depend on different  $g_{kk}$  and  $h_{kl}$ . A central limit theorem would require to derive their joint distribution, but this is not available because they are computed on different, partially overlapping data sets.

## 5. Monte Carlo simulation

In the previous section some of the CholCov's theoretical and asymptotic properties were discussed. In this simulation study we investigate its properties on realistic samples in a small bivariate setting and a larger problem of dimension 20. We want to focus on the use of the Cholesky decomposition to ensure positive semidefiniteness. As such we only use pre-averaging estimators, and compare our sequential estimation with full and (PSD-corrected) pairwise estimation. We consider the CholCov, as well as the MRC and Composite MRC. The web Appendix reports simulations which show similar results for the equivalent Kernel-based estimators.

### 5.1. Setup

As in Barndorff-Nielsen et al. (2011), we generate hypothetical prices, with  $Y^{(k)}(s)$  the associated log-price of asset  $k$ , from the log-price diffusion given by

$$\begin{aligned} dY_t^{(k)} &= \mu^{(k)}ds + dV_t^{(k)} + dF_t^{(k)} \\ dV_t^{(k)} &= \rho^{(k)} \sigma_t^{(k)} dB_t^{(k)} \\ dF_t^{(k)} &= \sqrt{1 - (\rho^{(k)})^2} \sigma_t^{(k)} dW_t, \end{aligned}$$



with  $k = 1, \dots, d$ . All  $B^{(k)}$  as well as  $W$  are independent Brownian motions.  $F^{(k)}$  denotes the common factor, scaled by  $\sqrt{1 - \rho^2}$  to determine its strength.

Each  $Y^{(k)}$  is a diffusive SV model with drift  $\mu^{(k)}$ . Their random spot volatility are given by  $\sigma^{(k)} = \exp(\beta_0^{(k)} + \beta_1^{(i)} \varrho^{(k)})$ , with  $d\varrho^{(k)} = \alpha^{(k)} \varrho^{(k)} dt + dB^{(k)}$ . The correlation between the changes in  $Y^{(k)}$  and  $Y^{(l)}$  is constant and equals  $\sqrt{1 - (\rho^{(k)})^2} \sqrt{1 - (\rho^{(l)})^2}$ .

We calibrate the parameters  $(\mu, \beta_0, \beta_1, \alpha, \rho)$  at  $(0.03, -5/16, 1/8, -1/40, -0.3)$  as in Barndorff-Nielsen et al. (2011). The stationary distribution of  $\varrho$  is utilized to restart the process each day at  $\varrho(0) \sim N(0, (-2(\beta)^2/\alpha)^{-1})$ . The parameter choice ensures that  $E\left(\int_0^1 \sigma^{(k)2}(u) du\right) = 1$ . The fact that  $\rho$  is set equal for all  $i$  leads to an equicorrelation structure with common correlation coefficient 0.91.

Microstructure noise is added to the return log-prices as  $X^{(k)} = Y^{(k)} + \epsilon^{(k)}$  with

$$\epsilon^{(k)} \mid \sigma, X \stackrel{iid}{\sim} N(0, \omega^2) \quad \text{with} \quad \omega^2 = \xi^2 \sqrt{N^{-1} \sum_{j=1}^N \sigma^{(k)4}(j/N)}.$$

Hence, the variance of the noise increases with the variance of the underlying process, in line with evidence from Bandi and Russell (2006).

Finally, independent Poisson processes are used to extract irregular, non-synchronous data from the complete high-frequency dataset. Each Poisson process is governed by a parameter  $\lambda^{(k)}$ , resulting in on average one observation every  $\lambda^{(k)}$  seconds for series  $k$ . On average, the series are observed  $23,400/\lambda^{(k)}$  times.

For each of the estimators, the bias and RMSE for variance and covariance elements are computed separately. The bias and RMSE, for a given element  $(k, l)$  of the matrix,  $\hat{\Sigma}_{k,l}$  are defined as

$$\text{Bias}_{k,l} = \hat{\Sigma}_{k,l} - \Sigma_{k,l} \quad \text{and} \quad \text{RMSE}_{k,l} = \sqrt{(\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2}. \quad (17)$$

We report the averages of the above two statistics over  $S = 1000$  replications. For the large dimensional case (i.e.  $d = 20$ ), we consider average bias and RMSE over multiple elements. Additionally, to get one general measure of accuracy, the average of the Frobenius distances is reported. For any d-dimensional matrix, the Frobenius distance is defined as

$$\text{Frobenius Distance} = \sum_{1 \leq k, l \leq d} (\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2. \quad (18)$$

## 5.2. Simulation Results

*Simulation I.* In the first simulation we consider only two assets, to determine a base level of performance. As a benchmark, we consider the Modulated Realized Covariance (MRC) of Christensen et al. (2010). For details of its implementation, see Appendix B. The estimates in the CholCov are similarly based on pre-averaging methods. We report the bias and RMSE with respect to the ICov for both estimators in Table 1, separating results for each individual element in the different panels. For each of the three noise-to-signal ratios  $\xi^2$ , we present the results using different Poisson process parameters  $\lambda$ , governing the liquidity of the stocks.

First consider the variances in Panel C, for the most liquid stock, and Panel D for the less liquid stock. In the CholCov, they are estimated univariately, and thus utilize all observations. For the MRC,

they are the result of the multivariate estimation on the common grid. The variance estimates thus differ in the number of observations used, and the optimized individual bandwidth for the CholCov.

These differences are evident in the results. The CholCov variance estimates typically have lower or similar bias as the MRC variance estimates. Not surprisingly, for the two CholCov variances, the estimate for the liquid stock has much lower bias and RMSE than that of the illiquid series. The bias of the MRC estimates is somewhere in between the bias of the two CholCov variances, but it has much higher RMSE across the board. The MRC performs poorly in the highest noise case, where its bias and RMSE are much higher than the CholCov's.

The covariance element is reported in Panel A. Here the table shows that the CholCov has slightly higher bias, but again, much lower RMSE. For both estimators, the bias of the covariance is in the same direction as the variances, suggesting the bias will be diminished for the correlations, reported in Panel B. Indeed, the correlations are generally close to their true value, and are estimated quite precisely with low RMSE. The only case that stands out is again the high noise case for the MRC. Here the bias in the variances is of larger magnitude than the bias in the covariance, and as such there is a relatively large downward bias in the correlations.

Overall, the simulation suggests that in the  $2 \times 2$  case our estimator has very similar performance to the MRC. Using the same class of estimators, the CholCov generally has slightly higher bias, but much lower RMSE.

*Simulation II.* The main advantages of the CholCov should be achieved in larger dimensional matrices with heterogeneity in trading intensity. To illustrate, here we simulate  $d = 20$  series with  $\lambda = \{5, 5, \dots, 5, 120\}$ .<sup>4</sup> The first nineteen series are observed on average every five seconds, and the last series once every two minutes. The MRC on the full dataset will generally use less than  $23,400/120$  observations, even though for the vast majority of the series more data is available. Our estimators will use the high observation frequency of the liquid series for all elements but those involving the last series.

We consider the MRC as well as its Composite counterpart, denoted cMRC. For the Composite estimator, as in Lunde et al. (2012), we first estimate the variances,  $D$ , using the univariate version of the estimator, after which realized correlations,  $R$ , are estimated on pairs of data. Similar to the CholCov, these estimators not only have the advantage of better data-sampling, but also optimal bandwidth selection for each element. The resulting estimate of the covariance matrix,  $DRD$ , will not necessarily be PSD, so any possible negative eigenvalues are set to zero as in Barndorff-Nielsen and Shephard (2004b). Their performance is compared in terms of the three aforementioned criteria. For the bias and RMSE the results are split up by reporting the averages of those elements involving the illiquid stock, and those that do not separately.

Table 2 reports the results of the simulation. Panel A reports general statistics for the full matrix. The CholCov and MRC estimators are both PSD by construction, and there is a large difference in average Frobenius Distance (FD) with respect to the ICOV, with the MRC's value being about five times as high. The pairwise estimator, cMRC, has a FD that is in line with the CholCov, but slightly

---

<sup>4</sup>Unreported simulations show similar, if less pronounced, conclusions with  $\lambda = 5$  for all assets and  $\lambda = \{2, 4, \dots, 38, 40\}$ .

Table 1: Results Simulation I

|                       | CholCov                               |       | MRC   |       | CholCov                                |       | MRC    |       |
|-----------------------|---------------------------------------|-------|-------|-------|--|-------|--------|-------|
|                       | Bias                                  | RMSE  | Bias  | RMSE  | Bias                                   | RMSE  | Bias   | RMSE  |
|                       | <i>Panel A: Integrated Covariance</i> |       |       |       | <i>Panel B: Integrated Correlation</i> |       |        |       |
| $\xi^2 = 0.000$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.010                                | 0.086 | 0.004 | 0.149 | -0.009                                 | 0.021 | -0.001 | 0.020 |
| $\lambda = (5, 10)$   | -0.013                                | 0.097 | 0.008 | 0.165 | -0.010                                 | 0.023 | -0.001 | 0.023 |
| $\lambda = (10, 20)$  | -0.017                                | 0.114 | 0.010 | 0.190 | -0.012                                 | 0.027 | -0.002 | 0.026 |
| $\lambda = (30, 60)$  | -0.033                                | 0.149 | 0.019 | 0.244 | -0.014                                 | 0.035 | -0.005 | 0.034 |
| $\lambda = (60, 120)$ | -0.049                                | 0.182 | 0.039 | 0.287 | -0.016                                 | 0.042 | -0.007 | 0.038 |
| $\xi^2 = 0.001$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.010                                | 0.087 | 0.004 | 0.150 | -0.010                                 | 0.022 | -0.003 | 0.021 |
| $\lambda = (5, 10)$   | -0.014                                | 0.099 | 0.005 | 0.165 | -0.011                                 | 0.024 | -0.003 | 0.023 |
| $\lambda = (10, 20)$  | -0.019                                | 0.115 | 0.010 | 0.192 | -0.012                                 | 0.028 | -0.005 | 0.027 |
| $\lambda = (30, 60)$  | -0.033                                | 0.149 | 0.022 | 0.242 | -0.015                                 | 0.037 | -0.008 | 0.034 |
| $\lambda = (60, 120)$ | -0.052                                | 0.182 | 0.035 | 0.284 | -0.017                                 | 0.046 | -0.012 | 0.041 |
| $\xi^2 = 0.010$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.015                                | 0.095 | 0.005 | 0.151 | -0.013                                 | 0.032 | -0.023 | 0.029 |
| $\lambda = (5, 10)$   | -0.019                                | 0.107 | 0.005 | 0.165 | -0.017                                 | 0.035 | -0.026 | 0.032 |
| $\lambda = (10, 20)$  | -0.023                                | 0.125 | 0.008 | 0.191 | -0.017                                 | 0.041 | -0.031 | 0.038 |
| $\lambda = (30, 60)$  | -0.033                                | 0.165 | 0.024 | 0.246 | -0.016                                 | 0.057 | -0.041 | 0.051 |
| $\lambda = (60, 120)$ | -0.049                                | 0.194 | 0.032 | 0.286 | -0.021                                 | 0.075 | -0.050 | 0.061 |
|                       | <i>Panel C: Integrated Variance 1</i> |       |       |       | <i>Panel D: Integrated Variance 2</i>  |       |        |       |
| $\xi^2 = 0.000$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.002                                | 0.077 | 0.005 | 0.159 | -0.006                                 | 0.091 | 0.002  | 0.155 |
| $\lambda = (5, 10)$   | -0.005                                | 0.086 | 0.010 | 0.177 | -0.009                                 | 0.102 | 0.006  | 0.171 |
| $\lambda = (10, 20)$  | -0.009                                | 0.103 | 0.013 | 0.204 | -0.011                                 | 0.122 | 0.008  | 0.197 |
| $\lambda = (30, 60)$  | -0.018                                | 0.137 | 0.026 | 0.263 | -0.035                                 | 0.161 | 0.017  | 0.248 |
| $\lambda = (60, 120)$ | -0.030                                | 0.169 | 0.049 | 0.309 | -0.058                                 | 0.199 | 0.039  | 0.296 |
| $\xi^2 = 0.001$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.003                                | 0.078 | 0.007 | 0.160 | -0.005                                 | 0.092 | 0.005  | 0.156 |
| $\lambda = (5, 10)$   | -0.005                                | 0.088 | 0.009 | 0.176 | -0.009                                 | 0.104 | 0.006  | 0.171 |
| $\lambda = (10, 20)$  | -0.009                                | 0.103 | 0.016 | 0.205 | -0.015                                 | 0.125 | 0.011  | 0.198 |
| $\lambda = (30, 60)$  | -0.019                                | 0.136 | 0.031 | 0.260 | -0.034                                 | 0.165 | 0.025  | 0.250 |
| $\lambda = (60, 120)$ | -0.032                                | 0.171 | 0.048 | 0.308 | -0.059                                 | 0.198 | 0.040  | 0.291 |
| $\xi^2 = 0.010$       |                                       |       |       |       |  |       |        |       |
| $\lambda = (3, 6)$    | -0.004                                | 0.086 | 0.029 | 0.162 | -0.008                                 | 0.100 | 0.025  | 0.158 |
| $\lambda = (5, 10)$   | -0.003                                | 0.097 | 0.034 | 0.177 | -0.010                                 | 0.116 | 0.029  | 0.173 |
| $\lambda = (10, 20)$  | -0.009                                | 0.113 | 0.041 | 0.207 | -0.015                                 | 0.135 | 0.033  | 0.199 |
| $\lambda = (30, 60)$  | -0.014                                | 0.152 | 0.069 | 0.269 | -0.031                                 | 0.183 | 0.060  | 0.253 |
| $\lambda = (60, 120)$ | -0.028                                | 0.182 | 0.083 | 0.314 | -0.045                                 | 0.215 | 0.078  | 0.298 |

*Note:* Simulation results of the multivariate factor diffusion with  $d = 2$ . The different panels report average Bias and RMSE over 1000 simulations for the individual elements of the covariance matrix.

lower in the low noise situations. Moreover, the estimate is often non-PSD. Its corrected alternative, denoted PSD cMRC, has similar Frobenius distance.

The variance estimates for the CholCov and cMRC are identical, so the difference in the correlation matrix, is of particular interest. The FD between the estimated and the true correlation matrix shows a slightly different picture than that of the covariance matrix. Where the PSD-corrected cMRC's FD had uniformly lower values, for the covariance matrix, the CholCov has lower FD in the presence of noise for the correlations. The difference becomes very large for  $\xi^2 = 0.010$  where the cMRC's FD is 50% larger.

Panel B reports more details on the individual elements. First consider the MRC. As everything is estimated on a common grid, there is no real difference in the bias and RMSE for each of the elements between the liquid and illiquid stocks. Compared to the other estimators, even if the bias is sometimes similar, the RMSE is large for all elements, explaining the large FD.

The CholCov and cMRC are much more accurate. The only problem occurs in the variance

Table 2: Results Simulation II

| Panel A                               | CholCov |  | MRC    |  | cMRC  |  | PSD cMRC |  |
|---------------------------------------|---------|--|--------|--|-------|--|----------|--|
| <i>Frobenius Distance Covariance</i>  |         |  |        |  |       |  |          |  |
| $\xi^2 = 0.000$                       | 5.855   |  | 30.932 |  | 5.162 |  | 5.161    |  |
| $\xi^2 = 0.001$                       | 5.663   |  | 31.555 |  | 5.227 |  | 5.224    |  |
| $\xi^2 = 0.010$                       | 5.819   |  | 32.809 |  | 5.846 |  | 5.808    |  |
| <i>Frobenius Distance Correlation</i> |         |  |        |  |       |  |          |  |
| $\xi^2 = 0.000$                       | 0.179   |  | 0.620  |  | 0.168 |  | 0.170    |  |
| $\xi^2 = 0.001$                       | 0.166   |  | 0.667  |  | 0.179 |  | 0.181    |  |
| $\xi^2 = 0.010$                       | 0.206   |  | 0.843  |  | 0.293 |  | 0.304    |  |
| <i>Panel B: Fraction PSD</i>          |         |  |        |  |       |  |          |  |
| $\xi^2 = 0.000$                       | 1.000   |  | 1.000  |  | 0.377 |  | 1.000    |  |
| $\xi^2 = 0.001$                       | 1.000   |  | 1.000  |  | 0.199 |  | 1.000    |  |
| $\xi^2 = 0.010$                       | 1.000   |  | 1.000  |  | 0.000 |  | 1.000    |  |

| Panel B                     | CholCov |       | MRC    |       | cMRC   |       | PSD cMRC |       |
|-----------------------------|---------|-------|--------|-------|--------|-------|----------|-------|
|                             | Bias    | RMSE  | Bias   | RMSE  | Bias   | RMSE  | Bias     | RMSE  |
| <i>Covariance Liquid</i>    |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.002  | 0.086 | 0.026  | 0.230 | -0.001 | 0.090 | -0.001   | 0.090 |
| $\xi^2 = 0.001$             | -0.002  | 0.088 | 0.029  | 0.233 | -0.001 | 0.091 | -0.001   | 0.091 |
| $\xi^2 = 0.010$             | -0.003  | 0.092 | 0.035  | 0.247 | -0.003 | 0.098 | -0.003   | 0.098 |
| <i>Covariance Illiquid</i>  |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.035  | 0.138 | 0.025  | 0.226 | -0.035 | 0.143 | -0.035   | 0.143 |
| $\xi^2 = 0.001$             | -0.037  | 0.136 | 0.029  | 0.230 | -0.034 | 0.145 | -0.034   | 0.144 |
| $\xi^2 = 0.010$             | -0.040  | 0.144 | 0.034  | 0.241 | -0.034 | 0.157 | -0.034   | 0.154 |
| <i>Correlation Liquid</i>   |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.001  | 0.018 | -0.004 | 0.037 | 0.000  | 0.017 | -0.001   | 0.017 |
| $\xi^2 = 0.001$             | -0.001  | 0.018 | -0.004 | 0.039 | -0.001 | 0.018 | -0.001   | 0.018 |
| $\xi^2 = 0.010$             | -0.001  | 0.021 | -0.008 | 0.060 | -0.002 | 0.026 | -0.007   | 0.026 |
| <i>Correlation Illiquid</i> |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.016  | 0.031 | -0.004 | 0.035 | -0.004 | 0.035 | -0.006   | 0.034 |
| $\xi^2 = 0.001$             | -0.001  | 0.029 | -0.004 | 0.036 | -0.004 | 0.036 | -0.008   | 0.035 |
| $\xi^2 = 0.010$             | 0.001   | 0.031 | -0.007 | 0.056 | -0.007 | 0.056 | -0.032   | 0.053 |
| <i>Variance Liquid</i>      |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.001  | 0.097 | 0.029  | 0.249 | -0.001 | 0.097 | -0.001   | 0.097 |
| $\xi^2 = 0.001$             | -0.001  | 0.098 | 0.033  | 0.253 | -0.001 | 0.098 | -0.001   | 0.098 |
| $\xi^2 = 0.010$             | -0.002  | 0.109 | 0.046  | 0.275 | -0.002 | 0.109 | 0.004    | 0.109 |
| <i>Variance Illiquid</i>    |         |       |        |       |        |       |          |       |
| $\xi^2 = 0.000$             | -0.058  | 0.199 | 0.026  | 0.227 | -0.058 | 0.199 | -0.055   | 0.198 |
| $\xi^2 = 0.001$             | -0.055  | 0.204 | 0.033  | 0.232 | -0.055 | 0.204 | -0.049   | 0.202 |
| $\xi^2 = 0.010$             | -0.048  | 0.216 | 0.042  | 0.245 | -0.048 | 0.216 | -0.009   | 0.207 |

*Note:* Simulation results of the multivariate factor diffusion with  $d = 20$  with  $\lambda = 5, 5, \dots, 5, 120$ . Panel A reports statistics for the full matrix. Panel B reports the average bias and RMSE of the covariance, correlation and variances, displayed separately for those elements involving the illiquid asset and those that do not.

estimate of the illiquid series, which is common both estimators. This transfers to a bias in the covariances involving the illiquid assets, but the correlation is estimated accurately. On average, the CholCov estimates the correlations with the least bias and the lowest RMSE, amongst the estimators considered.

The simulation demonstrates that in a large dimensional setting, the sequential estimation of the integrated correlation utilizing the Cholesky decomposition offers vast improvements over full-dimensional estimation, and slight improvements in the Frobenius Distance over pairwise estimation. However, the CholCov is PSD by construction, whereas the Composite estimator needs a rotation to become PSD.

## 6. Empirical Illustration I: Value-at-Risk forecasting

We expect our estimator to be especially useful in realistic large-scale portfolio applications that often require the estimator to be positive semidefinite and invertible. In our application we consider the forecasting of portfolio Value-at-Risk (VaR). When computing a portfolio VaR one has the option to either model the portfolio univariately or multivariately. For the univariate approach one uses the weights to compute portfolio returns and estimate its VaR based on the single series. Alternatively, one could estimate and model the full covariance matrix, and determine the portfolio VaR based on the multivariate setting. This has advantages for several reasons. First one can immediately calculate risk estimates for many different portfolios. Additionally, it has the advantage that it can be used for dynamic portfolio allocation, such as for instance a minimum variance portfolio. Finally, and most importantly, the dynamics of each of the volatility and correlation components are modeled separately. Santos et al. (2013) argue that for large dimensions, the information due to the multivariate modeling outweighs the additional uncertainty of estimating many parameters, and leads to better forecasts.

As such, here we take the latter approach and study the efficiency gains of using intraday data for VaR forecast accuracy. We compare forecasts from models estimated on daily returns with estimates based on intradaily techniques, i.e. dynamic models applied to CholCov estimates. We also consider the RCOV estimated on 5 minute previous-tick returns as a base case for using high-frequency estimators. Giot and Laurent (2004) and Brownlees and Gallo (2010) make the comparison between daily and high-frequency measures in a univariate setting. We are unaware of any paper comparing daily and intradaily models in the multivariate Value-at-Risk setting. For simplicity, in this application, we consider just two types of portfolios, equally- and value-weighted.

### 6.1. Data

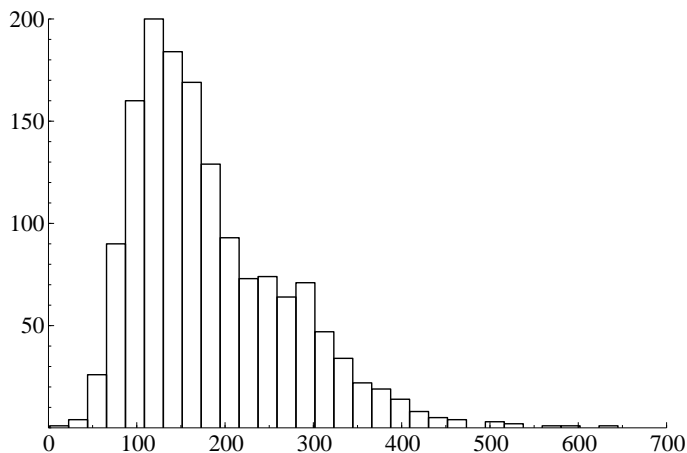
We analyze the portfolio risk for a total of 52 of the largest U.S. financial institutions.<sup>5</sup> We obtain trade data from January 2007 till December 2012 for a total of 1499 observation days. We clean the data using the step-by-step cleaning procedure of Barndorff-Nielsen et al. (2009). This entails the following rules: (P1) Delete all entries with a time-stamp outside the 9:30-16:00 exchange opening hours. (P2) Delete all entries with zero or negative price. (P3) Retain entries originating from a single exchange. For 43 firms this is NYSE, for the remaining 9 this is NASDAQ. (T1) Delete all corrected trades, where  $CORR \neq 0$ . (T2) Delete observations with an abnormal sale condition, i.e.  $COND$  has a letter other than 'E' or 'F'. (T3) Multiple transactions at the same time stamp are combined into a single observation at the median price. (T4\*) Delete an observation when the price deviated by more than 10 mean absolute deviations from a rolling centered median of 50 observations.

Throughout the analysis we use open-to-close returns obtained from the TAQ data. We use open-to-close as our ICov estimators do not take into account the overnight return. If we were to include overnight returns, gains in estimation accuracy would be more difficult to distinguish, as the overnight return is relatively dominant.

---

<sup>5</sup>The Tickers are: ACAS, AET, AFL, AIG, AIZ, ALL, AMP, AXP, BAC, BBT, BEN, BK, BLK, BRKB, CB, CBG, CINF, CMA, COF, CVH, EV, FITB, FNF, GNW, GS, HBAN, HIG, HNT, ICE, JNS, KEY, MET, MTB, NTRS, NYX, PFG, PGR, PNC, PRU, RF, SEIC, SNV, STI, STT, TMK, TROW, UNH, UNM, USB, WFC, WU, ZION.

Figure 3: Frequency plot of the daily number of observations after refresh-time synchronization of the 52 data series.



The weights for the value-weighted portfolio are proportional to firms' market capitalization, determined by the shares outstanding (from CRSP) times the closing price of the stock.

The estimation problem is moderately large with 52 firms and synchronization of the data will greatly reduce the total number of observations. In Figure 3 we plot the frequencies of number of observations after refresh-time synchronization of all the series. The least amount of observations is 6 while the most is 624. The median is 161. There are 11 days where the number of observations is smaller than the dimension of the problem. This invalidates the use of the traditional full-dimension estimators like the MRC and MKernel, which are no longer invertible. Second, our application, like many applications involving covariance estimates, requires the estimate to be positive semidefinite, invalidating the composite estimation technique. As such, the only estimator that is guaranteed to be positive semidefinite - without an arbitrary correction - and uses efficient sampling is ours.

## 6.2. Methodology

Our aim is to forecast portfolio Value-at-Risk. For a given  $d$ -dimensional vector of weights  $w_t$ , we compute the portfolio VaR over the open-to-close horizon on day  $t$  as follows:

$$VaR_t^q = w_t' \mu_{t|t-1} + z_q \sqrt{w_t' H_{t|t-1} w_t}, \quad (19)$$

where  $\mu_{t|t-1}$  is the vector conditional means and  $H_{t|t-1}$  the conditional covariance matrix. Under the assumption of conditional normality,  $z_q$  is the  $q$  quantile of the standard normal distribution. The conditional normality assumption is hard to justify for single stocks, but the dynamic quantile test that we will perform does not reject it in our portfolio setting.

We consider the Value-at-Risk of both long and short positions, setting  $q = \{0.01, 0.025, 0.05, 0.95, 0.975, 0.99\}$ . Long positions consider the left tail of the distribution, whereas for short positions the right tail is important. The Value-at-Risk is for long positions when  $q < 0.5$  and for short positions when  $q > 0.5$ .

The conditional mean is forecasted using AR(p) models, where the optimal lag order is individually determined by means of the Schwarz Information Criterion.

The CholCov and RCOV are ex-post measures, while for estimating the VaR we need a covariance forecast. As such we need a dynamic model to forecast the ICov. We consider two types of models. In the first we impose the same dynamics on all elements, leading to a Scalar-BEKK type specification, like the HEAVY model of Noreldin et al. (2012). In the second model that we consider, we allow for separate dynamics for the individual volatilities and correlations, like in the cRDCC of Bauwens et al. (2012). The performance of these two modeling strategies is compared with their counterparts using only daily returns, i.e. the Scalar-BEKK (Engle and Kroner, 1995) and cDCC (Aielli, 2013), respectively.

Before presenting the results, let us first detail the different estimation methods. The Scalar-BEKK / HEAVY models with dependence parameters  $\alpha$  and  $\beta$  take the form of

$$H_{t|t-1} = (1 - \alpha - \beta)\Omega + \alpha V_{t-1} + \beta H_{t-1|t-2}, \quad (20)$$

where for the Scalar-BEKK  $V_t = \varepsilon_t \varepsilon_t'$ , with  $\varepsilon_t$  corresponding to the vector of demeaned returns, and for the HEAVY model  $V_t = \hat{\Sigma}_t$ , the ICov estimate based on intraday data. To reduce the number of parameters to be estimated, we apply covariance targeting, where  $\Omega$  is the unconditional variance-covariance matrix of daily returns for the Scalar-BEKK and the average CholCov/RCOV for the HEAVY model. The HEAVY model additionally has a correction term, to match the unconditional variance of the model to that of daily returns. We implement the version of Equation (11) in Noreldin et al. (2012).

The cDCC models take the following form:

$$\begin{aligned} H_{t|t-1} &= D_{t|t-1} R_{t|t-1} D_{t|t-1} \\ R_{t|t-1} &= \text{diag}(Q_{t|t-1})^{-1/2} Q_{t|t-1} \text{diag}(Q_{t|t-1})^{-1/2} \\ Q_{t|t-1} &= (1 - \alpha - \beta)\bar{Q} + \alpha P_{t|t-1}^* + \beta Q_{t-1|t-2}, \end{aligned} \quad (21)$$

where  $P_t^* = \text{diag}(Q_{t|t-1})^{1/2} D_{t|t-1}^{-1} V_t D_{t|t-1}^{-1} \text{diag}(Q_{t|t-1})^{1/2}$ ,  $V_t$  is defined as above, i.e.  $V_t = \varepsilon_t \varepsilon_t'$  for the cDCC and  $V_t = \hat{\Sigma}_t$  for the cRDCC. Note that to reduce the number of parameters to be estimated, we do correlation targeting, by replacing  $\bar{Q}$  by the mean of  $P_t^*$ . Both cDCC models can be estimated in two steps, where first univariate models are fitted to estimate the volatilities  $D_t$ , which are then used to estimate the conditional correlation  $R_t$ . We have to specify the univariate models. There is ample evidence (e.g. Andersen et al., 2003) for the presence of long memory in realized variances, and the separation of volatilities and correlations allows us to model them as such. We use ARFIMA(1,  $d$ , 0) models on the natural logarithm of the estimated variances. By modeling the natural logarithm of the variances, we ensure positive out-of-sample forecasts.<sup>6</sup> For a fair comparison we model the volatilities in the cDCC on daily returns using a long memory model as well, i.e. a FIGARCH(1,  $d$ , 1).

The Scalar-BEKK and cDCC are estimated using a Composite Gaussian Likelihood, while the HEAVY and cRDCC models are estimated using a Composite Wishart Likelihood, all on contiguous pairs. Composite likelihood techniques for large-dimensional ARCH-type models were developed in Engle et al. (2008). They facilitate estimation, and reduce bias in parameter estimates present in

---

<sup>6</sup>We adjust for the bias caused by the transformation as in Giot and Laurent (2004).

large-dimensional problems. We assume Gaussian innovations for the computation of the VaR for all models.

To obtain the forecasts, we estimate all the models on an increasing window of observations, making one-step-ahead forecasts for the 1000 last days, re-estimating the parameters daily.

We test the out-of-sample performance of the VaR estimates in two ways. The first method is the dynamic quantile test of Engle and Manganelli (2004). They define a *Hit* variable associated with the ex-post observation of a VaR violation at time  $t$ :

$$Hit_t(q) = \begin{cases} 1 - q & \text{if } w'r_t < VaR_{t|t-1}^q \\ -q & \text{otherwise} \end{cases} \quad (22)$$

Similarly,  $Hit_t(1-q) = 1-q$  if  $w'r_t > VaR_{t|t-1}^{1-q}$ . We run the regressions  $Hit_t(q) = \delta + \sum_{k=1}^K \beta_k Hit_{t-k}(q) + \varepsilon_t$  and test the joint hypothesis  $H_0 : \delta = \beta_1 = \dots = \beta_K = 0, \forall k = 1, \dots, K$ . VaR violations are uncorrelated over time if the  $\beta$  are 0, whereas the unconditional coverage is correct if  $\delta = 0$ . Denote by  $\Psi = (\delta, \beta_1, \dots, \beta_K)'$  the vector of parameters of the model and by  $X$  the matrix of explanatory variables of the regression. The test statistic is

$$\frac{\hat{\Psi}' X' X \hat{\Psi}}{q(1-q)}, \quad (23)$$

and follows a  $\chi^2$ -distribution with  $K + 1$  degrees of freedom under the null of correct specification. Second, as in Chen and Gerlach (2013), we measure performance by means of a loss function and the Model Confidence Set of Hansen et al. (2011). They suggest the following loss function applicable to quantile forecasts, which is the criterion function minimized in quantile regressions:

$$LF_t^q = (VaR_{t|t-1}^q - w'r_t) Hit_t(q). \quad (24)$$

The forecasts lead to series of the loss function for each model and each quantile. The Model Confidence Set (MCS) allows statistical comparison of the competing models by comparing their respective loss series. It uses a block-bootstrap to obtain the distribution of the mean loss and eliminates models based on the maximum of pairwise t-statistics, automatically controlling the familywise error. Models are removed from the set until the final model set includes the best models with 90% confidence.<sup>7</sup>

### 6.3. Results

Table 3 reports the results of the Dynamic Quantile Test. The first three columns depict the results for the long-positions, and the last three columns show the results for the short positions. The top panel reports the p-values for the Equal Weighted (EW) portfolio and the bottom panel shows the results for the Value Weighted (VW) portfolio. Each panel contains successively the results for the two models using daily returns, Scalar-BEKK and cDCC, and the two models utilizing the CholCov, HEAVY and cRDCC. We give the results for  $K = 1$  and  $K = 2$ , but they are qualitatively similar for larger  $K$ . The results are available upon request.

---

<sup>7</sup>The MCS is carried out using the MulCom package for Ox, using 10,000 resamples and a bootstrap block-length of 2.



First consider the models on daily returns, the Scalar-BEKK and cDCC. The empirical results given in both panels tell the same story. The models using just daily returns are not flexible enough to accurately forecast the VaR. The Scalar-BEKK's unconditional coverage is rejected in many cases, and the rejection across the board shows that the violations are also dependent. The cDCC does perform a lot better, taking into account possible long-memory in volatility, but fails to model the left tail of the distribution adequately, with the p-values for the null hypothesis of the different versions of the test often smaller than 0.05, when  $q = 1\%$  or  $q = 2.5\%$ . We have considered alternative specifications, not reported for brevity, which included leverage effects, but these were also rejected. It is unclear whether the rejection is due to model misspecification or non-normal returns. As such we also estimated the model using the more flexible multivariate Student distribution. Again, this did not lead to significant improvements.

By increasing the information set to include intraday data, we can estimate the models on RCOV and CholCov. The HEAVY model on RCOV offers some slight improvements over the daily Scalar-BEKK. However, the model is still almost uniformly rejected. The model on CholCov offers some more significant improvements. However, the short memory HEAVY model is still rejected by the data. For the CholCov, based on  $K = 1$  the returns might be conditionally Gaussian, but rejection for  $K = 2$  shows misspecification. This suggests we need more lags, and as such we have considered a HEAVY(2,2) model, which is rejected by the data in a similar manner.

The cRDCC with ARFIMA dynamics on the variances takes into account the long-memory properties of Realized Variance. The model on RCOV has similar performance to the cDCC on daily returns, with rejections in the left tail. However, the flexibility of the model, combined with the accurate CholCov estimates, allows it to capture both the unconditional coverage and pass the test for independence of violations, with only a single rejection at the 5 percent level for the 1% VaR at  $K = 2$ .

To see the economic significance of using liquidity sorting instead of observation count sorting, Table 3 also reports the results using the CholCov based on observation count sorting along with the empirically best model, the ARFIMA cRDCC. The resulting VaR forecasts are inferior. For the Value Weighted portfolio, rejections are similar to the liquidity sorting based forecasts, but for the Equal Weighted portfolio there are again problems in the left tail, with rejections for the two most extreme quantiles.

Second we consider the evaluation of the loss function on VaR forecasts. Table 4 shows which models are part of the 90% Model Confidence Set for each VaR quantile and for both portfolios. The cRDCC model on CholCov is always in the MCS for both portfolios. Nearer the centre of the distribution, the cDCC on daily returns and cRDCC on RCOV are also included in the MCS. For the right tail of the return distribution, the CholCov using observation count sorting are part of the MCS. In general, the Scalar-BEKK / HEAVY -type models are rejected here as well.

The results suggest that, for diversified portfolios like the ones considered here, accurate VaR forecasts can be obtained under the assumption of conditional normality, by utilizing the information content of high-frequency data, an efficient PSD estimate of the covariance and an appropriate dynamic model.

Table 3: Dynamic Quantile Test P-Values

|   | K\q | 1.0%  | 2.5%  | 5.0%  | 95.0% | 97.5% | 99.0% |
|---|-----|-------|-------|-------|-------|-------|-------|
| Equal Weighted                                    |     |       |       |       |       |       |       |
| Scalar-BEKK                                       | 1   | 0.018 | 0.000 | 0.011 | 0.001 | 0.000 | 0.000 |
|   | 2   | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| cDCC  | 1   | 0.029 | 0.015 | 0.335 | 0.499 | 0.754 | 0.950 |
|   | 2   | 0.022 | 0.034 | 0.227 | 0.566 | 0.779 | 0.976 |
| HEAVY-RCOV  | 1   | 0.000 | 0.001 | 0.102 | 0.198 | 0.016 | 0.000 |
|   | 2   | 0.000 | 0.003 | 0.152 | 0.000 | 0.000 | 0.000 |
| cRDCC-RCOV  | 1   | 0.028 | 0.270 | 0.764 | 0.902 | 0.591 | 0.042 |
|   | 2   | 0.022 | 0.219 | 0.548 | 0.012 | 0.002 | 0.078 |
| HEAVY-CholCov                                     | 1   | 0.385 | 0.204 | 0.557 | 0.262 | 0.800 | 0.042 |
|   | 2   | 0.000 | 0.007 | 0.015 | 0.000 | 0.000 | 0.011 |
| cRDCC-CholCov<br><i>Observation Count Sorting</i> | 1   | 0.010 | 0.008 | 0.272 | 0.777 | 0.899 | 0.746 |
|   | 2   | 0.002 | 0.014 | 0.210 | 0.213 | 0.944 | 0.856 |
| cRDCC-CholCov<br><i>Liquidity Sorting</i>         | 1   | 0.746 | 0.719 | 0.531 | 0.688 | 0.413 | 0.797 |
|   | 2   | 0.078 | 0.831 | 0.240 | 0.414 | 0.547 | 0.918 |
| Value Weighted                                    |     |       |       |       |       |       |       |
| Scalar-BEKK                                       | 1   | 0.000 | 0.000 | 0.002 | 0.011 | 0.006 | 0.000 |
|   | 2   | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| cDCC  | 1   | 0.130 | 0.030 | 0.133 | 0.506 | 0.667 | 0.236 |
|   | 2   | 0.055 | 0.057 | 0.246 | 0.712 | 0.735 | 0.354 |
| HEAVY-RCOV  | 1   | 0.887 | 0.014 | 0.002 | 0.000 | 0.000 | 0.000 |
|   | 2   | 0.060 | 0.028 | 0.004 | 0.000 | 0.000 | 0.000 |
| cRDCC-RCOV  | 1   | 0.065 | 0.041 | 0.914 | 0.674 | 0.474 | 0.036 |
|   | 2   | 0.000 | 0.076 | 0.954 | 0.174 | 0.003 | 0.068 |
| HEAVY-CholCov                                     | 1   | 0.565 | 0.021 | 0.283 | 0.380 | 0.838 | 0.385 |
|   | 2   | 0.000 | 0.003 | 0.189 | 0.000 | 0.000 | 0.085 |
| cRDCC-CholCov<br><i>Observation Count Sorting</i> | 1   | 0.385 | 0.193 | 0.135 | 0.788 | 0.719 | 0.887 |
|   | 2   | 0.000 | 0.214 | 0.253 | 0.253 | 0.831 | 0.944 |
| cRDCC-CholCov<br><i>Liquidity Sorting</i>         | 1   | 0.950 | 0.675 | 0.900 | 0.946 | 0.667 | 0.625 |
|   | 2   | 0.039 | 0.315 | 0.920 | 0.387 | 0.735 | 0.809 |

*Note:* P-Values of the Dynamic Quantile Test for the Equal Weighted and Value Weighted portfolio Value-at-Risks. The Scalar-BEKK and cDCC are estimated on daily returns, while the last two models are estimated using the proposed CholCov estimate for the Integrated Covariance.

Table 4: Loss function MCS results

| q  | 1.0% | 2.5% | 5.0% | 95.0% | 97.5% | 99% |
|--|------|------|------|-------|-------|-----|
| Equal Weighted                                 |      |      |      |       |       |     |
| Scalar-BEKK                                    |      |      |      |       |       |     |
| cDCC   |      |      |      | ✓     | ✓     | ✓   |
| HEAVY-RCOV                                     |      |      |      |       |       |     |
| cRDCC-RCOV                                     |      | ✓    | ✓    | ✓     |       |     |
| HEAVY-CholCov                                  |      |      |      |       | ✓     | ✓   |
| cRDCC-CholCov <i>Observation Count Sorting</i> |      |      |      | ✓     | ✓     | ✓   |
| cRDCC-CholCov <i>Liquidity Sorting</i>         | ✓    | ✓    | ✓    | ✓     | ✓     | ✓   |
| Value Weighted                                 |      |      |      |       |       |     |
| Scalar-BEKK                                    |      |      |      |       |       |     |
| cDCC   |      |      |      | ✓     | ✓     |     |
| HEAVY-RCOV                                     |      |      |      |       |       |     |
| cRDCC-RCOV                                     |      |      | ✓    | ✓     |       |     |
| HEAVY-CholCov                                  |      |      | ✓    |       |       |     |
| cRDCC-CholCov <i>Observation Count Sorting</i> |      |      |      |       | ✓     | ✓   |
| cRDCC-CholCov <i>Liquidity Sorting</i>         | ✓    | ✓    | ✓    | ✓     | ✓     | ✓   |

*Note:* The table shows the models included in the 90% Model Confidence set for the different VaR forecast quantiles and the two portfolios, based on the loss function (24).

## 7. Empirical Illustration II: Forecasting betas

A further application in which we expect the CholCov to be of practical relevance is the forecasting of dynamic exposures to observed risk factors, which we will refer to as betas. As in the previous section, this requires an accurate and positive semidefinite estimator of the covariance matrix. We estimate the exposures of the 52 financial institutions in our previous exercise to ten sector Exchange Traded Funds (ETFs), being SPY, XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, XLY. Our sample for the ETFs only runs until April 30th 2012, for a total of 1336 observations.

We compare the betas produced by one-step-ahead forecasts of the cDCC-type models for CholCov, RCOV and daily returns. As our sample is slightly reduced, we now consider the last 800 observations our out-of-sample window. The betas are obtained by forecasting the 11-dimensional covariance matrix, of the 10 ETFs and the individual financial institution. We then obtain the 10-dimensional beta vector as

$$\hat{\beta}_{k,t|t-1} = \hat{\Sigma}_{kE,t|t-1} \hat{\Sigma}_{EE,t|t-1}^{-1},$$

where the covariance matrix is partitioned into the firm  $k$  and the ten ETFs  $E$ :

$$\Sigma_{t|t-1} = \begin{bmatrix} \Sigma_{kk,t|t-1} & \Sigma_{Ek,t|t-1} \\ \Sigma_{kE,t|t-1} & \Sigma_{EE,t|t-1} \end{bmatrix}.$$

We use a regression-based comparison that was proposed by Engle (2012). As such, we then define the variables

$$Z_{k,t|t-1} = \hat{\beta}'_{k,t|t-1} ETF_t,$$

where  $ETF_t$  is the vector of realized sector returns. We consider three different estimators of  $\hat{\beta}_k$ ,

based on the CholCov, RCOV and daily return, leading to three different auxiliary variables, denoted  $Z_{k,t|t-1}^{CholCov}$ ,  $Z_{k,t|t-1}^{RCOV}$ , and  $Z_{k,t|t-1}^{DCC}$ . We then estimate the following model by OLS:

$$r_{k,t} = \alpha + \delta_k^{CholCov} Z_{k,t|t-1}^{CholCov} + \delta_k^{RCOV} Z_{k,t|t-1}^{RCOV} + \delta_k^{DCC} Z_{k,t|t-1}^{DCC} + \eta_{k,t}, \quad (25)$$

for each  $k = 1, \dots, 52$ . A perfect specification for the  $\beta$ -vector would lead to a  $\delta$  coefficient equal to one, while the other coefficients would be zero. Engle (2012) do an in-sample comparison of a static beta with a DCC beta, and Hansen et al. (2013) add the Realized Beta GARCH to the analysis and perform an out-of-sample analysis.

We can test which model is 'superior' by testing the following three null hypotheses:

$$\begin{aligned} H_{CholCov} : \quad & \delta_k^{CholCov} = 1, \quad \delta_k^{RCOV} = \delta_k^{DCC} = 0 \\ H_{RCOV} : \quad & \delta_k^{RCOV} = 1, \quad \delta_k^{CholCov} = \delta_k^{DCC} = 0 \\ H_{DCC} : \quad & \delta_k^{DCC} = 1, \quad \delta_k^{CholCov} = \delta_k^{RCOV} = 0, \end{aligned}$$

taking into account heteroskedasticity by using White (1980) standard errors. Additionally, we wish to identify the benefits of using the CholCov over the RCOV. To test this, we consider the regression (25) but without  $Z_{k,t|t-1}^{DCC}$ , and test the hypotheses:

$$\begin{aligned} H'_{CholCov} : \quad & \delta_k^{CholCov} = 1, \quad \delta_k^{RCOV} = 0 \\ H'_{RCOV} : \quad & \delta_k^{RCOV} = 1, \quad \delta_k^{CholCov} = 0 \end{aligned}$$

Second, we compare the betas in terms of a beta hedging tracking exercise. We compute the time series  $(r_{k,t} - Z_{k,t|t-1}^{CholCov})$ ,  $(r_{k,t} - Z_{k,t|t-1}^{RCOV})$  and  $(r_{k,t} - Z_{k,t|t-1}^{DCC})$  for each asset  $k$ , and seek the series with the smallest sample variance by means of the Model Confidence Set of Hansen et al. (2011).

The results are reported in Table 5. The first two panels show the results from the regression based test, whereas Panel C shows the beta hedging exercise. Panel A shows descriptive statistics of the cross-sectional variation of the estimated  $\delta$  coefficients across the 52 assets. The beta stemming from the CholCov estimates has the highest weight on average and cannot be rejected as the 'superior' model in over 40 percent of the assets. It almost completely encompasses the explanatory power of beta forecasts based on the RCOV, which obtains very low loadings and is always rejected. The model based on daily returns has slightly higher loadings, but is also almost always rejected.

In the direct comparison between the CholCov and RCOV forecasts, the CholCov provides superior results, and cannot be rejected in almost 60 percent of the cases.

The beta hedging exercise provides the same conclusion. We obtain Model Confidence Sets, which contain the set of models with the smallest tracking error with a probability of no less than 90%. The CholCov tracking portfolios are never rejected, and the DCC based portfolios are in slightly over 40% of the sets.

## 8. Conclusions

We propose an ex-post estimator of the integrated covariance that uses the Cholesky decomposition to obtain an estimate that is ensured positive semidefinite. The elements are estimated sequentially, on an increasing set of series. As such, the estimator uses many more observations than the traditional multivariate estimators, but fewer than pairwise estimators. The Cholcov is flexible and can use

Table 5: Beta Comparisons

| Panel A: Summary of parameter estimates |                      |                   |                  |                      |                   |                      |                   |                  |
|---|----------------------|-------------------|------------------|----------------------|-------------------|----------------------|-------------------|------------------|
|   | Full Model           |                   |                  | Excluding DCC        |                   | Single series        |                   |                  |
|   | $\delta_i^{CholCov}$ | $\delta_i^{RCOV}$ | $\delta_i^{DCC}$ | $\delta_i^{CholCov}$ | $\delta_i^{RCOV}$ | $\delta_i^{CholCov}$ | $\delta_i^{RCOV}$ | $\delta_i^{DCC}$ |
| Mean                                    | 0.724                | 0.064             | 0.258            | 0.922                | 0.122             | 0.976                | 0.995             | 1.017            |
| Stdev.                                  | 0.199                | 0.142             | 0.230            | 0.131                | 0.151             | 0.103                | 0.325             | 0.086            |
| 5%                                      | 0.461                | -0.178            | -0.083           | 0.702                | -0.102            | 0.841                | 0.394             | 0.880            |
| Median                                  | 0.713                | 0.077             | 0.268            | 0.912                | 0.138             | 0.957                | 0.997             | 1.018            |
| 95%                                     | 1.061                | 0.293             | 0.637            | 1.115                | 0.368             | 1.117                | 1.457             | 1.137            |

| Panel B: Rejection Frequencies |                    |                   |                        |                     |
|--------------------------------|--------------------|-------------------|------------------------|---------------------|
| $H_{CholCov} : 0.577$          | $H_{RCOV} : 1.000$ | $H_{DCC} : 0.962$ | $H'_{CholCov} : 0.404$ | $H'_{RCOV} : 1.000$ |

| Panel C: Frequency in Model Confidence Set |             |            |
|--|-------------|------------|
| CholCov: 1.000                             | RCOV: 0.038 | DCC: 0.442 |

*Note:* Comparison of the betas obtained from CholCov and RCOV cRDCC forecasts, and cDCC forecasts based on daily returns. Panel A presents descriptive statistics of the estimated coefficients of regression (25). Panel B shows the rejection frequencies for the hypothesis outlined in the text and Panel C reports the frequency that each of the models are included in the MCS in the beta hedging exercise.

any other estimator for the intermediate calculations, adopting both their robustness and asymptotic properties. Simulations confirm this and show that the resulting estimates are accurate.

In an empirical application we use the CholCov in a portfolio setting which requires the covariance matrix estimate to be Positive Semidefinite. The problem is moderately large, involving over 50 stocks. Using an appropriate dynamic model, which allows for long memory in the variances, we forecast portfolio Value-at-Risk and are unable to reject the model, using standard normal quantiles. This in contrast to models based on daily returns and dynamic models on the CholCov not allowing for long memory, which are rejected by the data.

As a second application we forecast conditional betas with respect to ten sector ETFs. We evaluate them using a regression-based comparison and a beta hedging tracking exercise. We find that the forecasts based on the CholCov provide the best results in terms of both evaluation criteria, compared to forecasts based on RCOV and daily returns.

## 9. References

- Aielli, G. P., 2013. Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* (Forthcoming).
- Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105 (492), 1504–1517.
- Andersen, T., Bollerslev, T., Diebold, F., 2007. Roughing it up: including jump components in the measurement, modelling and forecasting of return volatility. *The Review of Economics and Statistics* 89, 701–720.
- Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Wu, G., 2006. Realized beta: Persistence and predictability. *Advances in Econometrics* 20, 1–39.
- Bandi, F., Russell, J., 2006. Separating microstructure noise from volatility. *Journal of Financial Economics* 79 (3), 655–692.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., 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., 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, 149–169.
- Barndorff-Nielsen, O. E., Shephard, N., 2004a. Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.
- Barndorff-Nielsen, O. E., Shephard, N., 2004b. Measuring the impact of jumps in multivariate price processes using bipower covariation. Working Paper.
- Bauwens, L., Storti, G., Violante, F., 2012. Dynamic conditional correlation models for realized covariance matrices. Working Paper.
- Bos, C. S., Janus, P., Koopman, S. J., 2012. Spot variance path estimation and its application to high-frequency jump testing. *Journal of Financial Econometrics* 10 (2), 354–389.
- Boudt, K., Cornelissen, J., Croux, C., 2012. Jump robust daily covariance estimation by disentangling variance and correlation components. *Computational Statistics and Data Analysis* 56, 2993–3005.
- Boudt, K., Croux, C., Laurent, S., 2011. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance* 18, 353–367.
- Boudt, K., Zhang, J., 2014. Jump robust two time scale covariance estimation and realized volatility budgets. *Quantitative Finance*, Forthcoming.
- Brownlees, C. T., Gallo, G. M., 2010. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* 8 (1), 29–56.
- Chen, Q., Gerlach, R. H., 2013. The two-sided weibull distribution and forecasting financial tail risk. *International Journal of Forecasting* 29 (4), 527–540.
- Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26 (6), 922–947.
- Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* 159, 116–133.
- Doornik, J., 2009. Object-oriented Matrix Programming Using Ox. Timberlake Consultants Press.
- Engle, R., 2012. Dynamic conditional beta. Working Paper.
- Engle, R., Shephard, N., Sheppard, K., 2008. Fitting vast dimensional time-varying covariance models. Working Paper.
- Engle, R. F., 2002. Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R. F., Kroner, K. F., 1995. Multivariate simultaneous generalized ARCH. *Econometric theory* 11 (01), 122–150.
- Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22 (4), 367–381.
- Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* 107 (497), 412–428.
- Giot, P., Laurent, S., 2004. Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance* 11 (3), 379–398.
- Hansen, P. R., Lunde, A., 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24 (2), 127–161.
- Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. *Econometrica* 79 (2), 453–497.
- Hansen, P. R., Lunde, A., Voev, V., 2013. Realized beta garch: Multivariate garch model with realized measures of volatility and covolatility. *Journal of Applied Econometrics* (Forthcoming).
- Harris, F., McInish, T., Shoesmith, G., Wood, R., 1995. Cointegration, error correction, and price discovery on informationally linked security markets. *Journal of Financial and Quantitative Analysis* 30, 563–581.
- Hautsch, N., Kyj, L., Oomen, R., 2012. A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics* 27, 627–645.
- Hautsch, N., Podolskij, M., 2013. Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. *Journal of Business & Economic Statistics* 31 (2), 165–183.
- Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.
- Hounyo, U., Gonçalves, S., Meddahi, N., 2013. Bootstrapping pre-averaged realized volatility under market microstructure noise. Working Paper.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications* 119 (7), 2249–2276.
- Kristensen, D., 2010. Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory* 26 (01), 60–93.
- Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix for stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10, 603–621.
- Lunde, A., Shephard, N., Sheppard, K., 2012. Econometric analysis of vast covariance matrices using composite realized kernels. Working Paper.
- Miller, K. S., 1981. On the inverse of the sum of matrices. *Mathematics Magazine* 54 (2), pp. 67–72.
- Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 27 (6), 907–933.

- Palandri, A., 2009. Sequential conditional correlations: Inference and evaluation. *Journal of Econometrics* 153, 122–132.
- Pelletier, D., 2006. Regime switching for dynamic correlations. *Journal of Econometrics* 131 (1), 445–473.
- Peluso, S., Corsi, F., Mira, A., 2012. A bayesian high-frequency estimator of the multivariate covariance of noisy and asynchronous returns. Working Paper.
- Santos, A. A., Nogales, F. J., Ruiz, E., 2013. Comparing univariate and multivariate models to forecast portfolio value-at-risk. *Journal of Financial Econometrics* 11 (2), 400–441.
- Schmelzer, T., Hauser, R., 2013. Seven sins in portfolio optimization. Working Paper.
- Shephard, N., Xiu, D., 2013. Econometric analysis of multivariate realised qml: estimation of the covariation of equity prices under asynchronous trading. Working Paper.
- Tang, G., 1995. Intertemporal stability in international stock market relationships: A revisit. *The Quarterly Review of Economics and Finance* 35, 579–593.
- Tsay, R., 2010. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics Series. Wiley.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.
- Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160, 33–47.

## Appendix A. Notation

Indices  $i, j$  are used for time.  $k, l, m, n$  are used for the  $d$  different assets.

| Refresh time  |   |
|---|---|
| Number of assets  | $d$   |
| Set of assets   | $\mathcal{D}$   |
| Calendar time grid (e.g. seconds)                               | $t_i$   |
| Refresh-time grid on series $\mathcal{d} \subseteq \mathcal{D}$ | $\tau_i^{\mathcal{d}}$                                  |
| Number of observations refresh sample                           | $N^{\mathcal{d}}$                                       |
| Number of refresh-times of constant correlation                 | $\kappa$  |
| Decompositions  |   |
| Spot covariance matrix  | $\Sigma(s)$   |
| Spot volatility   | $\sigma_k(s) = \Sigma_{k,k}^{1/2}(s)$                   |
| Spot correlation  | $\rho_{k,l}(s)$   |
| Diagonal matrix with spot volatility                            | $D(s)$  |
| Spot correlation matrix   | $R(s)$  |
| Lower diagonal matrix Cholesky                                  | $H(s)$ with elements $h_{k,l}$                          |
| Diagonal matrix Cholesky  | $G(s)$ with elements $g_{k,l}$                          |
| Series  |   |
| Price observations  | $X_{\tau_j}$  |
| Returns   | $r_{\tau_j}^{(1)}, \dots, r_{\tau_j}^{(d)}$             |
| Cholesky projections  | $f_{\tau_j}^{(1)}, \dots, f_{\tau_j}^{(d)}$             |
| Devolitized returns   | $\hat{u}_{\tau_j}^{(1)}, \dots, \hat{u}_{\tau_j}^{(d)}$ |

## Appendix B. Implementation Details

*Pre-Averaged Estimators.* The building block of pre-averaged estimators are the pre-averaged returns defined as

$$\bar{r}_{\tau_j}^{(k)} = \sum_{h=1}^{k_N-1} g\left(\frac{h}{k_N}\right) r_{\tau_j+h}^{(k)}.$$

The function  $g : [0, 1] \rightarrow \mathbb{R}$  is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative  $g'$  with  $g(0) = g(1) = 0$  satisfying  $\int_0^1 g^2(s) ds > 0$ . Define the following functions

and numbers associated with  $g$ :

$$\begin{aligned}\phi_1(s) &= \int_s^1 g'(u)g'(u-s)du, \\ \phi_2(s) &= \int_s^1 g(u)g(u-s)du, \\ \psi_1 &= \phi_1(0), \quad \psi_2 = \phi_2(0), \\ \Phi_{11} &= \int_0^1 \phi_1^2(s)ds, \quad \Phi_{12} = \int_0^1 \phi_1(s)\phi_2(s)ds, \quad \Phi_{22} = \int_0^1 \phi_2^2(s)ds.\end{aligned}$$

The functions  $\phi_1$  and  $\phi_2$  are assumed 0 outside the interval  $[0,1]$ . Finite sample estimators of these quantities are available in Christensen et al. (2010) and will be denoted with a superscript  $k_N$ . Following Hautsch and Podolskij (2013); Christensen et al. (2010) we use  $g(x) = \min(x, 1-x)$ .

$k_N$  is a sequence of integers satisfying  $k_N = \lfloor \theta_u N^{1/2} \rfloor$ . We use  $\theta_u = 0.8$  for the univariate estimator during our entire paper based on simulations and recommendations in Hautsch and Podolskij (2013). The pre-averaged returns are simply a weighted average over the returns in a local window. This averaging diminishes the influence of the noise. The order of the window size  $k_N$  is chosen to lead to optimal convergence rates. The pre-averaging estimator is then simply the analogue of the Realized Variance but based on pre-averaged returns and an additional term to remove bias due to noise, which we denote Modulated Realized Variance (MRV):

$$MRV = \frac{N^{-1/2}}{\theta \psi_2^{k_N}} \sum_{i=0}^{N-k_N+1} (\bar{r}_{\tau_i})^2 - \frac{\psi_1^{k_N} N^{-1}}{2\theta^2 \psi_2^{k_N}} \sum_{i=0}^N r_{\tau_i}^2. \quad (\text{B.1})$$

The multivariate counterpart is very similar and was proposed in Christensen et al. (2010). The estimator is called the Modulated Realized Covariance (MRC) and is defined as

$$MRC = \frac{N}{N - k_N + 2} \frac{1}{\psi_2^{k_N} k_N} \sum_{i=0}^{N-k_N+1} \bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_i' - \frac{\psi_1^{k_N}}{\theta^2 \psi_2^{k_N}} \hat{\Psi}, \quad (\text{B.2})$$

where  $\bar{\mathbf{r}}_i = (\bar{r}_i^{(1)} \ \bar{r}_i^{(2)} \ \dots \ \bar{r}_i^{(d)})$  denotes the matrix of all pre-averaged return series and  $\hat{\Psi}_N = \frac{1}{2N} \sum_{i=1}^N \mathbf{r}_{\tau_i} (\mathbf{r}_{\tau_i})'$ . The second term is a bias correction to make it consistent. However, due to this correction the estimator is not ensured PSD. An alternative is to slightly enlarge de bandwidth such that  $k_N = \lfloor \theta_m N^{1/2+\delta} \rfloor$ .  $\delta = 0.1$  results in a consistent estimate without the bias correction and a PSD estimate, in which case:

$$MRC^\delta = \frac{N}{N - k_N + 2} \frac{1}{\psi_2^{k_N} k_N} \sum_{i=0}^{N-k_N+1} \bar{\mathbf{r}}_{\tau_i} \cdot \bar{\mathbf{r}}_{\tau_i}'.$$

In this paper we use  $\theta_m = 1.0$  for the multivariate version and slightly increase the bandwidth with  $\delta = 0.1$  to get a consistent estimator. For the CholCov we use the normal bandwidth and apply the bias correction.

### Appendix C. CholCov asymptotics using pre-averaging estimators

The Modulated Realized Variance (MRV), defined in Appendix B, is essentially a realized variance on pre-averaged returns, with a number of bias corrections. As our estimate  $g_{kk}$  is exactly this estimator on transformed series, its asymptotic distribution follows directly from that of the MRV.

**Proposition 1** *Under assumptions  $\mathbf{H}$  and  $\mathbf{K}$  in Jacod et al. (2009), the estimates of  $g_{kk}$  conditional*



on no estimation error in previous estimates converge stably in law as

$$N^{1/4} (\hat{g}_{kk} - g_{kk}) \xrightarrow{d_s} N(0, \Gamma),$$

where  $\Gamma = \frac{4}{\psi_2^2} \left( \phi_{22} \theta \sigma^{*4} + 2\phi_{12} \frac{\sigma^{*2} \Psi^*}{\theta} + \phi_{11} \frac{\Psi^{*2}}{\theta^3} \right)$ ,  $\Psi^*$  is the variance of the noise process of the transformed series and  $d_s$  denotes stable convergence in distribution.

*Proof:* see Theorem 3.1 in Jacod et al. (2009).

Similarly, the  $h_{lk}$  estimates follow the distribution of the realized beta based on the multivariate pre-averaging estimator, the Modulated Realized Covariance (MRC).

**Proposition 2** Assume that  $\mathbb{E}(|\epsilon^j|^8) < \infty$  for all  $j = 1, \dots, d$ . Define  $\nabla = \hat{\Sigma}_{kk}^{-1} (1, -\hat{h}_{lk})$ , where  $\hat{\Sigma}_{kk}$  is the  $kk$ -th element of the MRC estimate. As  $N \rightarrow \infty$  the distribution of  $h_{lk}$  conditional on all previous estimates is as follows

$$N^{1/4} (\hat{h}_{lk} - h_{lk}) \xrightarrow{d} MN(0, \nabla \bar{\Gamma} \nabla'), \quad (\text{C.1})$$

with

$$\bar{\Gamma} = \begin{bmatrix} \widehat{avar}_{(k-1)d+l, (k-1)d+l}^* & \widehat{avar}_{(k-1)d+l, (k-1)d+k}^* \\ \bullet & \widehat{avar}_{(k-1)d+k, (k-1)d+k}^* \end{bmatrix}, \quad (\text{C.2})$$

where

$$avar_{MRC} = \frac{2}{\psi_2^2} \left( \Phi_{22} \theta \int_0^1 \Lambda(s) ds + \frac{\Phi_{12}}{\theta} \int_0^1 \Theta(s) ds + \frac{\Phi_{11}}{\theta^3} \Upsilon \right). \quad (\text{C.3})$$

$\Lambda, \Theta$  and  $\Upsilon$  are  $d \times d \times d \times d$  arrays with elements

$$\begin{aligned} \Lambda_s &= \{\Sigma_{kk'}(s) \Sigma_{ll'}(s) + \Sigma_{kl'}(s) \Sigma_{lk'}(s)\}_{k, k', l, l'=1, \dots, d} \\ \Theta_s &= \{\Sigma_{kk'}(s) \Phi_{ll'} + \Sigma_{kl'}(s) \Phi_{kl'} + \Sigma_{k'l}(s) \Phi_{kl'} + \Sigma_{ll'}(s) \Phi_{kk'}\}_{k, k', l, l'=1, \dots, d} \\ \Upsilon &= \{\Psi_{kk'} \Psi_{ll'} + \Psi_{kl'} \Psi_{lk'}\}_{k, k', l, l'=1, \dots, d}. \end{aligned}$$

$avar^*$  is as in Equation (C.3), but with  $\Sigma = \Sigma^*$  and  $\Psi = \Psi^*$ .

*Proof:* see Theorem 6 in Christensen et al. (2010).

Feasible versions of the asymptotic distributions are discussed at length in Jacod et al. (2009) and Christensen et al. (2010). When estimating the feasible versions, one does not have to worry about the transformations of the data, and the resulting  $\Sigma^*$  and  $\Psi^*$ . Indeed, the feasible asymptotic variance of the different parameters can be estimated directly by computing them using the transformed series as input. Note that different versions are available, dependent on the bandwidth choice. The results presented here are based on the slightly larger bandwidth used in this paper.

# Research Papers 2013



- 2013-40: Daniel Ventosa-Santaulària and Carlos Vladimir Rodríguez-Caballero: Polynomial Regressions and Nonsense Inference
- 2013-41: Diego Amaya, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez: Does Realized Skewness Predict the Cross-Section of Equity Returns?
- 2013-42: Torben G. Andersen and Oleg Bondarenko: Reflecting on the VPN Dispute
- 2013-43: Torben G. Andersen and Oleg Bondarenko: Assessing Measures of Order Flow Toxicity via Perfect Trade Classification
- 2013-44: Federico Carlini and Paolo Santucci de Magistris: On the identification of fractionally cointegrated VAR models with the F(d) condition
- 2013-45: Peter Christoffersen, Du Du and Redouane Elkamhi: Rare Disasters and Credit Market Puzzles
- 2013-46: Peter Christoffersen, Kris Jacobs, Xisong Jin and Hugues Langlois: Dynamic Diversification in Corporate Credit
- 2013-47: Peter Christoffersen, Mathieu Fournier and Kris Jacobs: The Factor Structure in Equity Options
- 2013-48: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs and Mehdi Karoui: Illiquidity Premia in the Equity Options Market
- 2013-49: Peter Christoffersen, Vihang R. Errunza, Kris Jacobs and Xisong Jin: Correlation Dynamics and International Diversification Benefits
- 2013-50: Georgios Effraimidis and Christian M. Dahl: Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure
- 2013-51: Mehmet Caner and Anders Bredahl Kock: Oracle Inequalities for Convex Loss Functions with Non-Linear Targets
- 2013-52: Torben G. Andersen, Oleg Bondarenko, Viktor Todorov and George Tauchen: The Fine Structure of Equity-Index Option Dynamics
- 2014-01: Manuel Lukas and Eric Hillebrand: Bagging Weak Predictors
- 2014-02: Barbara Annicchiarico, Anna Rita Bennato and Emilio Zanetti Chini: 150 Years of Italian CO2 Emissions and Economic Growth
- 2014-03: Paul Catani, Timo Teräsvirta and Meiqun Yin: A Lagrange Multiplier Test for Testing the Adequacy of the Constant Conditional Correlation GARCH Model
- 2014-04: Timo Teräsvirta and Yukai Yang: Linearity and Misspecification Tests for Vector Smooth Transition Regression Models
- 2014-05: Kris Boudt, Sébastien Laurent, Asger Lunde and Rogier Quaedyvlieg: Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity