Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity.
This Version: December 3, 2013

Kris Boudt[a,b], Sébastien Laurent[c], Asger Lunde[d,e], Rogier Quaedvlieg[f,*]

[a] *Department of Business, Vrije Universiteit Brussel, Belgium.*
[b] *VU University Amsterdam, Netherlands.*
[c] *Aix-Marseille Univ. (Aix-Marseille School of Economics), CNRS & EHESS, France*
[d] *CREATES, University of Aarhus, Denmark*
[e] *School of Economics and Management, Aarhus University, Denmark*
[f] *Department of Finance, Maastricht University, Netherlands*

Abstract

We propose an estimator of the ex-post covariation of log-returns under asynchronicity and microstructure noise. We use a positive semidefinite factorization of 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, robust to microstructure noise and asynchronicity. An extensive Monte Carlo study confirms good finite sample properties. In our application we show that for a portfolio of 52 financial institutions, Value-at-Risk forecasts obtained using dynamic models utilizing our high-frequency estimator are accurate, in contrast to those from an approach using daily returns, for which the violations are serially correlated.

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

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 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 estimators have sacrificed the latter (e.g. Hayashi and Yoshida, 2005; Zhang, 2011; Lunde et al., 2012).
Up until recently, estimation of realized covariance at ultra high frequencies was made difficult by

[☆] This is a preliminary version, results are subject to change. Please do not distribute.
This research was financially supported by a grant of the Dutch Organization for Scientific Research (NWO). 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.

[*] 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@iae-aix.com (Sébastien Laurent), alunde@asb.dk (Asger Lunde), r.quaedvlieg@maastrichtuniversity.nl (Rogier Quaedvlieg)

two empirical phenomena which induce biases in the estimates. First, the presence of market microstructure noise (e.g. bid-ask bounce), and second, non-synchronous trading. To avoid these biases, covariances were estimated using returns at moderate frequencies, for instance every 20 minutes. 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, 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 separates groups of liquid and illiquid assets, applies the multivariate realized kernel to each group, and combines these estimates into the full matrix. Aït-Sahalia et al. (2010) synchronize pairs and use the 'polarization result' to estimate the covariance bivariately, 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. 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. The efficiency of data usage therefore lies somewhere in between that of PSD $d-$dimensional estimation, and non-PSD pure bivariate estimation.

Each bivariate estimate on 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 estimators include the sub-sampling estimator (Zhang, 2011; Boudt and Zhang, 2012), pre-averaging estimator (Christensen et al., 2010), kernel estimator (Barndorff-Nielsen et al., 2011) and likelihood based method (Aït-Sahalia et al., 2010). 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 for 52 assets. We apply recent advances in the literature on 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 that utilize the CholCov in their estimation greatly improve both unconditional coverage and independence between Value-at-Risk violations.

The paper is structured as follows. In Section 2 we consider the theoretical setup and outline the decomposition. Section 3 first discusses practical issues in preparing the data for estimation, and then presents the algorithm on how to obtain 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 extensive Monte Carlo Simulations designed

to highlight the improvements due to more efficient data-sampling. Section 6 provides the empirical application on the forecasting of portfolio Value-at-Risk. Finally, Section 7 concludes.

2. Theoretical setup

Our aim is to accurately estimate the Integrated Covariance (ICov) matrix of a $d-$dimensional Brownian semimartingale process $Y = (Y^{(1)}, \ldots, Y^{(d)})'$. 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 $X = (X^{(1)}, \ldots, X^{(d)})'$. Let the set of all series be denoted $\mathcal{D}$, with subsets $\mathcal{A} \subseteq \mathcal{D}$. 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 $0 \leq t_1^{(k)} \leq t_2^{(k)} \leq \ldots \leq t_{N_{(k)}}^{(k)} \leq 1$. X is driven by the efficient log-price Y, a Brownian semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t^0), P^0)$:

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

where μ is a $d\times 1$ predictable locally bounded drift process, W is a d-dimensional vector of independent Brownian motions and σ 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:

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

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

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

with $\mathbf{\Psi}$ 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 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. For simplicity we only consider noise of the form in Equation (2).

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

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

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

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

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

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

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.

$$R(s) = H(s)G(s)H(s)', \tag{5}$$

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 $d = 3$:

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

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

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

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

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

3.1. Dealing with asynchronicity

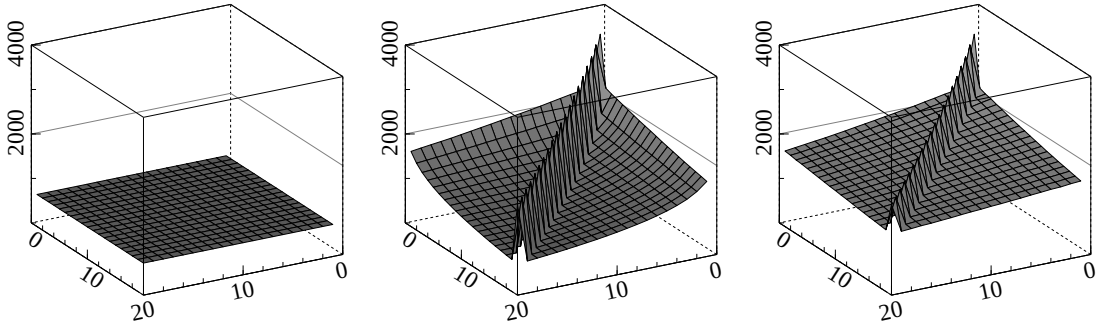
In high-frequency data, assets are traded at irregular intervals and seldom simultaneously. Moreover, not all stocks are equally liquid, such that the number of fresh prices within a fixed time interval may substantially differ across firms. One way to synchronize the data is by means of refresh-time 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. Barndorff-Nielsen et al. (2011) define refresh-time for $t \in [0,1]$ as follows. The first refresh-time occurs at $\tau_1 = \max(t_1^{(1)}, \ldots, t_1^{(d)})$, and subsequent refresh-times at $\tau_{j+1} = \max(t^{(1)}_{N^{(1)}_{\tau_j+1}}, \ldots, t^{(d)}_{N^{(d)}_{\tau_j+1}})$. The grid, denoted $\tau^{\mathscr{d}}$, is a function of the series over which they are sampled, with $\mathscr{d} \subseteq \mathcal{D}$. The grid's number of observations is denoted $N^{\mathscr{d}}$, while the individual series have size $N^{(k)}$. Finally, denote the durations as $\Delta_i^{\mathscr{d}} = \tau_i^{\mathscr{d}} - \tau_{i-1}^{\mathscr{d}}$. 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, $N^{\mathscr{d}} \leq \min_k N^{(k)}$, 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 (2004) 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

Figure 1: Number of observations used for each element in estimation of Σ



These graphs depict simulation based number of observations available for the estimation of elements in a 20×20 covariance matrix for different sampling schemes. The sampling frequencies follow independent poisson processes with parameter λ 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.

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 intensity 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 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:

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

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 actual observation times, or their difference Δ_i, 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.

3.3. Decomposition of the ICov assuming locally constant correlations

Consider the finest equispaced time grid t_i, with $\delta = t_{i+1} - t_i$ and $i = 1, ..., N_t$. 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, $\varkappa\delta$. Then, ICov can be approximated as follows:

$$\text{ICov} \approx \sum_{j=1}^{N_t/\varkappa} \left[\int_{t_{\varkappa(j-1)}}^{t_{\varkappa j}} D(s) R(t_{\varkappa(j-1)}) D(s) ds \right], \tag{10}$$

where $D(s)$ is the diagonal matrix whose kth element is the spot volatility of asset k and $R(t_{\varkappa(j-1)})$ is the constant correlation matrix between $\varkappa(j-1)$ and $\varkappa j$. Equation (10) is a discretized version of (3) and converges to ICov under some smoothness conditions when $N_t \to \infty, \varkappa \to 0$, but with $N_t/\varkappa \to \infty$.

In practise we are dealing with a refresh-time grid. We assume the spot correlation to be constant over a fixed number, $\varkappa$, of refresh-times. $\tau^{\mathcal{D}}$ is the refresh-time grid using all series under consideration, with $N^{\mathcal{D}}$ observations. $\varkappa$ is defined on this grid such that the spot correlation is constant for the $\varkappa$ observations at times $\tau^{\mathcal{D}}_{\varkappa(j-1)}, ..., \tau^{\mathcal{D}}_{\varkappa j}$ for $j = 1, ..., N^{\mathcal{D}}/\varkappa$. In the remainder of the paper, split-sample will refer to a sample partitioned into $N^{\mathcal{D}}/\varkappa$ samples based on transaction times $\tau^{\mathcal{D}}_{\varkappa j}$, regardless of what transaction grid the sample has. The number of observation within a subsample will generally be larger than $\varkappa$. When $N^{\mathcal{D}} \to \infty$, 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, $D(s)$ are estimated using a rolling window of length γ, as in for instance Boudt et al. (2012), and aggregated to the constant correlation windows. Asymptotic properties for a class of related estimators of local variance in absence of jumps can be found in Kristensen (2010). The correlations are estimated using the observations within the window $\tau_{\varkappa(j-1)}$

and $\tau_{\varkappa j}$. We then obtain

$$\text{ICov} \approx \sum_{j=1}^{N^{\mathcal{D}}/\varkappa} \left[\int_{\tau^{\mathcal{D}}_{\varkappa(j-1)}}^{\tau^{\mathcal{D}}_{\varkappa j}} D(s) R(\tau^{\mathcal{D}}_{\varkappa(j-1)}) D(s) ds \right]. \tag{11}$$

Finally, the question remains how to determine $\varkappa$ and γ 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.

For simplicity of presentation we will take $\varkappa = \gamma = 1/N^{\mathcal{D}}$, such that the constant correlation and variance periods equal one day, but asymptotically we require them to go to 0. The next section describes in detail how we estimate the correlation matrix R.

3.4. The CholCov

The first step is to estimate the variances and standardize the series. Define the series of intraday returns $r^{(1)}_{\tau_j}, \ldots, r^{(d)}_{\tau_j}$, where $r^{(k)}_{\tau_j} = X^{(k)}_{\tau_j} - X^{(k)}_{\tau_{j-1}}$. $D(s)$ can be estimated using any integrated variance estimator on the returns. Boudt et al. (2012) obtain the spot volatilities by estimating the Integrated Variance (IV) over rolling windows. An alternative would be to explicitly model the intraday periodicity (see e.g. Boudt et al., 2011). Next, data is pre-multiplied by $D(s)^{-1}$, which amounts to dividing all returns by their spot volatility. As the time intervals are irregularly spaced, the spot volatility, $\sigma^{(k)}(s)$, for the returns is proportional to the time interval between the two adjacent observations:

$$\hat{\sigma}^{(k)}_{\tau_j} = \sqrt{\frac{1}{N} \sum_{\tau_{j-1} \leq s \leq \tau_j} \hat{\sigma}^{2(k)}(s)}.$$

The standardized log-returns, are therefore

$$\hat{u}^{(k)}_{\tau_j} = \frac{r^{(k)}_{\tau_j}}{\hat{\sigma}^{(k)}_{\tau_j}}. \tag{12}$$

This ensures that the integrated variance of the series $\hat{u}^{(k)}_{\tau_j}$, $j = 1, ..., N^{\mathcal{D}}$, 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 in the estimation.

Next, the elements in the decomposition of the correlation matrix in (5) are estimated. By replacing R_{kk} and R_{kl} in the system of equations (7) and (8) with local estimates of the variance and covariance, we sequentially estimate the g_{kk} and h_{kl} terms. Each element requires only a subset of the data series. Tsay (2010) shows that the quantities in (7) and (8) are simply the coefficients and residual variances

of the orthogonal transformations

$$\begin{aligned} f_{\tau_j}^{(1)} &= \hat{u}_{\tau_j}^{(1)} \\ f_{\tau_j}^{(2)} &= \hat{u}_{\tau_j}^{(2)} - \hat{\beta}_{21} f_{\tau_j}^{(1)} \\ &\vdots \\ f_{\tau_j}^{(d)} &= \hat{u}_{\tau_j}^{(d)} - \sum_{l=1}^{d-1} \hat{\beta}_{dl} f_{\tau_j}^{(l)}, \end{aligned} \tag{13}$$

where $\hat{\beta}_{kl}$ are the least-squares coefficients in the regressions

$$\begin{aligned} \hat{u}_{\tau_j}^{(2)} &= \beta_{21} f_{\tau_j}^{(1)} \\ \hat{u}_{\tau_j}^{(3)} &= \beta_{31} f_{\tau_j}^{(1)} + \beta_{32} f_{\tau_j}^{(2)} \\ &\vdots \\ \hat{u}_{\tau_j}^{(d)} &= \sum_{l=1}^{d-1} \beta_{dl} f_{\tau_j}^{(l)}, \end{aligned} \tag{14}$$

where we left out the error terms in each regression for notational convenience. We therefore have $h_{kl} = \beta_{kl}$ and $g_{kk} = \mathrm{Var}(f_{\tau_j}^{(k)})$, with $f_{\tau_j}^{(k)} \perp\!\!\!\perp f_{\tau_j}^{(l)}$ for $k \neq l$. This system of equations may be estimated on a split-sample basis, again governed by $\tau^{\mathcal{D}}$ and $\varkappa$.

To put this in the context of the high-frequency estimator, we first consider a simple scenario in which we observe equispaced returns for all series, uncontaminated by microstructure noise. In this scenario Realized (Co)Variance is a consistent estimator of the Integrated (Co)Variance and OLS on intraday returns leads to consistent estimates of β. 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) (Andersen et al., 2003). The estimation then directly follows (13)-(14). g_{11} is obtained as the RV of the first return series. h_{21} is the OLS-coefficient from regressing $f^{(1)}$ on $\hat{u}^{(2)}$. g_{22} is the RV of the projected series $f^{(2)}$ from the previous regression. 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 next orthogonal transformation is the g_{33} element. This process continues until all elements have been estimated.

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 identical. The elements are computed sequentially, 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. We estimate the quantities using robust realized variances and realized betas[1] (Andersen et al., 2006). 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

[1]The realized beta is computed by estimating a 2×2 realized covariance matrix $\hat{\Sigma}$. The realized beta is then computed as $\hat{\beta}_{kl} = \hat{\Sigma}_{kl}/\hat{\Sigma}_{ll}$.

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.

Realized beta does not take into account dependence between regressors, and hence the estimator is only consistent if the orthogonal transformations $f^{(.)}$ 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 $f^{(.)}$ 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 likely to be the most accurate. The following algorithm summarizes the procedure:

CholCov Estimation Algorithm.

1. Apply refresh-time on $\mathscr{a} = \{1\}$ to obtain the grid $\tau^{\mathscr{a}}$. Estimate $\hat{g}_{11}$ using an IV estimator on $f^{(1)}_{\tau^{\mathscr{a}}_j} = \hat{u}^{(1)}_{\tau^{\mathscr{a}}_j}$.

2. Apply refresh-time on $\mathscr{b} = \{1, 2\}$ to obtain the grid $\tau^{\mathscr{b}}$. Estimate $\hat{\beta}^{\mathscr{b}}_{21}$ as the realized beta between $f^{(1)}_{\tau^{\mathscr{b}}_j}$ and $\hat{u}^{(2)}_{\tau^{\mathscr{b}}_j}$. Set $\hat{h}_{21} = \hat{\beta}^{\mathscr{b}}_{21}$. Estimate $\hat{g}_{22}$ using an IV estimator on $f^{(2)}_{\tau^{\mathscr{b}}_j} = \hat{u}^{(2)}_{\tau^{\mathscr{b}}_j} - \hat{h}_{21} f^{(1)}_{\tau^{\mathscr{b}}_j}$.

3. Apply refresh-time on $\mathscr{c} = \{1, 3\}$ to obtain the grid $\tau^{\mathscr{c}}$. Estimate $\hat{\beta}^{\mathscr{c}}_{31}$ as the realized beta between $f^{(1)}_{\tau^{\mathscr{c}}_j}$ and $\hat{u}^{(3)}_{\tau^{\mathscr{c}}_j}$. Set $\hat{h}_{31} = \hat{\beta}^{\mathscr{c}}_{31}$.

4. Apply refresh-time on $\mathscr{d} = \{1, 2, 3\}$ to obtain the grid $\tau^{\mathscr{d}}$. Re-estimate $\hat{\beta}^{\mathscr{d}}_{21}$ at the new grid, such that the projections are orthogonal. Estimate $\hat{\beta}^{\mathscr{d}}_{32}$ as the realized beta between $f^{(2)}_{\tau^{\mathscr{d}}_j} = \hat{u}^{(2)}_{\tau^{\mathscr{d}}_j} - \hat{\beta}^{\mathscr{d}}_{21} f^{(1)}_{\tau^{\mathscr{d}}_j}$ and $\hat{u}^{(3)}_{\tau^{\mathscr{d}}_j}$. Set $\hat{h}_{32} = \hat{\beta}^{\mathscr{d}}_{32}$. Estimate $\hat{g}_{33}$ using an IV estimator on $f^{(3)}_{\tau^{\mathscr{d}}_j} = \hat{u}^{(3)}_{\tau^{\mathscr{d}}_j} - \hat{h}_{32} f^{(2)}_{\tau^{\mathscr{d}}_j} - \hat{h}_{31} f^{(1)}_{\tau^{\mathscr{d}}_j}$.

5. Continue in the same fashion by sampling over $1, ..., k, l$ to estimate h_{lk} using the smallest possible set. Re-estimate the β_{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, $\hat{h}$.

The matrix $\hat{Q}(s) = \hat{H}(s)\hat{G}(s)\hat{H}(s)'$ is guaranteed to be a symmetric positive semidefinite matrix, but to qualify as a correlation matrix, it also needs that the diagonal elements are equal to one and that the off-diagonal elements are in the [-1,1] intervals. We considered two ways to ensure this. First, as in Equation (2.5) of Pelletier (2006) we could impose the constraint

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

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. The second approach is to simply standardize the matrix to unit diagonal, as in Engle (2002),

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

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

$$CholCov = \sum_{j=1}^{N^{\mathcal{D}}/\varkappa} \left[\sum_{\tau^{\mathcal{D}}_{\varkappa(j-1)} \leq s \leq \tau^{\mathcal{D}}_{\varkappa j}} \hat{D}(s)\hat{R}(\tau^{\mathcal{D}}_{\varkappa(j-1)})\hat{D}(s)(\tau_{\varkappa j} - \tau_{\varkappa(j-1)}) \right]. \tag{15}$$

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 (15) 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. Similarly we could use the jump robust two time scale covariance (Boudt and Zhang, 2012) and additionally obtain robustness against jumps. 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 (avar) of the estimates in $\hat{G}$ and $\hat{H}$. The consistency is not affected by the sequential estimation, but confidence bounds may be wider. 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 me 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 $\Sigma^* = \sigma^*(s)(\sigma^*(s))'$. A denotes a matrix containing 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

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

The asymptotic theory of the g and h elements follows directly from the theory of whichever estimator we use. In Appendix B we report for completeness these results when the pre-averaging estimator, as detailed in Appendix A, is used.

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. Since we want to focus on the PSD construction using the Cholesky decomposition and not the specific estimator for the elements, pre-averaging estimators are used throughout the simulations and the application. 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 $k = 1, \ldots, d$. 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 ϱ is utilized to restart the process each day at $\varrho(0) \sim N(0, (-2(\beta)^2/\alpha)^{-1})$. The parameter choice ensures that $\mathrm{E}\left(\int_0^1 \sigma^{(k)2}(u)du\right) = 1$. The fact that ρ 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

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

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

$$\text{Bias}_{k,l} = \hat{\Sigma}_{k,l} - \Sigma_{k,l} \quad \text{and RMSE}_{k,l} = \sqrt{(\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2}. \tag{16}$$

We report the averages over $S = 1000$ replications. For the large dimensional matrices ($d = 20$), 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

$$\text{Frobenius Distance} = \sum_{1 \le k,l \le d} (\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2. \tag{17}$$

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 A. 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.

Table 1 reports that in the 2×2 case, our estimator has similar performance to the MRC. In our simulations it has slightly higher bias, but much lower RMSE. In high noise cases there is a clear gain in terms of variance estimation, and thus correlation estimates. The simulation also illustrates the effect the number of observation has on estimation. The CholCov determines the variances separately, such that they are estimated at the highest possible observation frequency. The first series has on average twice the number of observations series two has, and the bias and RMSE of the variance for the liquid stock are much lower. This is indicative of the effect for larger dimensional estimates, and the gains we can achieve on the first estimates by sequential estimation.

Table 1: Results Simulation I

	CholCov		MRC		CholCov		MRC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	Panel A: Integrated Covariance				*Panel B: Integrated Correlation*			
$\xi^2 = 0.000$								
$\lambda = (3, 6)$	-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
	Panel C: Integrated Variance 1				*Panel D: Integrated Variance 2*			
$\xi^2 = 0.000$								
$\lambda = (3, 6)$	-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

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

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 $\lambda = \{5, 5, \ldots, 5, 120\}$[2]. 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

[2]Unreported simulations show similar, if less pronounced, conclusions with $\lambda = 5$ for all assets and $\lambda = \{2, 4, \ldots, 38, 40\}$.

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 (2004). 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. The CholCov and MRC estimators are both PSD by construction, and there is a large difference in average Frobenius distance, with the MRC's value being about five times as high. Looking at the details on the individual elements, this turns out to mostly be a variance issue, not an average bias issue. Regardless, the CholCov is very accurate, also in this large dimensional framework. The only real bias is in the variance of the illiquid assets, which transfers to the covariance. The cMRC has exactly the same problem, as the variance elements are estimated in the same way. However, with high levels of noise the cMRC is never PSD, and needs a correction. After the correction the cMRC still performs slightly better in terms of Frobenius Distance on the covariance matrix, but in the presence of noise the CholCov is more accurate for the correlations.

6. Empirical Illustration: 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. As such we compare forecasts from models estimated on daily returns with estimates based on intradaily techniques, i.e. dynamic models applied to CholCov estimates. Giot and Laurent (2004) and Brownlees and Gallo (2010) do this 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.[3] We obtain trade data from January 2007 till December 2012 for a total of 1499 observation days. We clean the

[3]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.

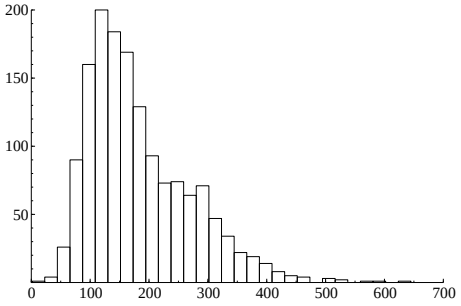
Table 2: Results Simulation II

Panel A	CholCov	MRC	cMRC	PSD cMRC
Frobenius Distance Covariance				
$\xi^2 = 0.000$	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
Frobenius Distance Correlation				
$\xi^2 = 0.000$	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
Panel B: Fraction PSD				
$\xi^2 = 0.000$	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
Covariance Liquid								
$\xi^2 = 0.000$	-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
Covariance Illiquid								
$\xi^2 = 0.000$	-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
Correlation Liquid								
$\xi^2 = 0.000$	-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
Correlation Illiquid								
$\xi^2 = 0.000$	-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
Variance Liquid								
$\xi^2 = 0.000$	-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
Variance Illiquid								
$\xi^2 = 0.000$	-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

Simulation results of the multivariate factor diffusion with $d = 20$ with $\lambda = 5, 5, \ldots, 5, 120$. 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.

Figure 2: Frequency plot of the sample's daily number of observations after synchronization.



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, we would no longer be able to test the accuracy of the realized estimators, as the overnight return is relatively dominant.

The weights for the value-weighted portfolio are proportional to firms' market capitalization, determined by 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 2 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 and uses efficient sampling is ours.

6.2. Methodology

Our aim is to forecast portfolio Value-at-Risk. Recall that for a given $d-$dimensional vector of weights w_t, the portfolio VaR equals

$$VaR_t^q = w_t'\mu_{t|t-1} + z_q\sqrt{w_t'H_{t|t-1}w_t}, \tag{18}$$

where $\mu_{t|t-1}$ is the vector conditional means, $H_{t|t-1}$ the conditional covariance matrix and z_q the q quantile of the standard normal distribution. The 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 $q = \{0.01, 0.025, 0.05\}$. Long positions consider the left tail of the distribution, whereas for short positions the right tail is important. The Value-at-Risk for long positions is denoted VaR^q, while for short positions we use VaR^{1-q}.
The conditional mean is forecasted using AR(p) models, where the optimal order is individually determined by means of the Schwarz Information Criterion.
The CholCov is an ex-post measure, while for estimating the VaR we need a covariance forecast. As such we need a dynamic model to forecast the ICov. We consider two model types: either we impose the same dynamics on all elements, leading to a Scalar-BEKK type specification, which was introduced as the HEAVY model by (Noureldin et al., 2012). Alternatively, we allow for separate dynamics for the individual volatilities and correlations, leading to the cRDCC (Bauwens et al., 2012). We will compare the performance of these two models with their counterparts using only daily returns, 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 take the form of

$$H_{t|t-1} = (1 - \alpha - \beta)\Omega + \alpha V_{t-1} + \beta H_{t-1|t-2} \tag{19}$$

where for the Scalar-BEKK $V_t = \varepsilon_t \varepsilon_t'$, with ε_t corresponding to the vector of demeaned returns, and for the HEAVY model $V_t = \hat{\Sigma}_t$, the CholCov estimate. To reduce the number of parameters to be estimated, we apply covariance targeting, where Ω is the unconditional variance-covariance matrix of returns for the Scalar-BEKK and the average CholCov for HEAVY. 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 Noureldin et al. (2012).
The cDCC models take the following form:

$$\begin{aligned} H_{t|t-1} &= D_{t|t-1} R_{t|t-1} D_{t|t-1} \\ R_{t|t-1} &= diag(Q_{t|t-1})^{-1/2} Q_{t|t-1} diag(Q_{t|t-1})^{-1/2} \\ Q_{t|t-1} &= (1 - \alpha - \beta)\bar{Q} + \alpha P^*_{t|t-1} + \beta Q_{t-1|t-2}, \end{aligned} \tag{20}$$

where $P^*_t = diag(Q_{t|t-1})^{1/2} D^{-1}_{t|t-1} V_t D^{-1}_{t|t-1} diag(Q_{t|t-1})^{1/2}$, where V_t is defined as before, with $V_t = \varepsilon_t \varepsilon_t'$ for the cDCC and $V_t = \hat{\Sigma}_t$ for the cRDCC. Again, there is correlation targeting, in the form of $\bar{Q}$, which is set to the mean of P^*_t. Both cDCC models can be estimated in two steps, where first univariate models are fitted to model the volatilities D_t, which are then used to model the correlation matrix R_t. A specification has to be chosen for 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 - to ensure positive out-of-sample forecasts - the natural logarithm of the CholCov variances.[4] For a fair comparison we model the volatilities in the cDCC on daily returns using a long memory model as

[4] We adjust for the bias caused by the transformation as in Giot and Laurent (2004).

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 parameters present in 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 using 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:

$$Hit_t(q) = \begin{cases} 1-q & \text{if } w'r_t < VaR^q_{t|t-1} \\ -q & otherwise. \end{cases} \tag{21}$$

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

$$\frac{\hat{\Psi}' X' X \hat{\Psi}}{q(1-q)}, \tag{22}$$

and follows a χ^2-distribution with $K+1$ degrees of freedom under the null of correct specification.

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 $K = 2$, but they are qualitatively similar for larger K. 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 for $k \geq 1$ 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 CholCov. The short memory HEAVY model is rejected by the data. Based on $k = 0$ or $k = 1$ 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. This allows the model to not only capture the unconditional coverage, but it also passes the test for independence of violations, with only a single rejection at the 5 percent level for the 1% VaR at $k = 2$.

The results show that we can obtain accurate forecasts of the covariance matrix in large portfolio settings, by utilizing the information content of high-frequency data, a PSD estimate of the covariance and an appropriate dynamic model.

7. 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 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 estimate to be Positive Semidefinite. The problem is moderately large with over 50 series considered. 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.

8. 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., 2004. Measuring the impact of jumps in multivariate price processes using bipower covariation. Working Paper.

Table 3: Dynamic Quantile Test P-Values

	k\q	1.0%	2.5%	5.0%	95.0%	97.5%	99.0%
				Equal Weighted			
	0	0.026	0.224	0.384	0.110	1.000	0.011
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
	0	0.011	0.005	0.147	0.310	0.685	1.000
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
	0	0.204	1.000	0.885	0.110	0.543	0.340
HEAVY	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
	0	0.525	1.000	0.310	0.384	0.311	0.525
cRDCC	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			
	0	0.057	0.224	0.468	0.147	0.839	0.001
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
	0	0.057	0.026	0.246	0.246	0.543	0.112
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
	0	0.340	0.543	0.885	0.192	0.839	0.204
HEAVY	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
	0	1.000	0.839	0.663	1.000	0.543	0.340
cRDCC	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

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.

Bauwens, L., Storti, G., Violante, F., 2012. Dynamic conditional correlation models for realized covariance matrices. Working Paper.
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., 2012. 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.
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., Shephard, N., Sheppard, K., 2008. Fitting vast dimensional time-varying covariance models. Discussion paper series n.403, Department of Economics, University of Oxford.
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.
Harris, F., McInisch, 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-diemnsional realized covariance estimation. Journal of Applied Econometrics 27, 627–645.
Hautsch, N., Podolskij, M., 2010. Pre-averaging based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. Discussion Paper 2010-038, CRC 649, Berlin.
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, 60–93.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix fo 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.
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.
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.
Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. Journal of Econometrics 160, 33–47.

Appendix A. Implementation Details

Pre-Averaged Estimators. In order to define the univariate pre-averaging estimator Hautsch and Podolskij (2010) we first define the pre-averaged returns as

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

The function $g : [0,1] \to \mathbb{R}$ 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 ϕ_1 and ϕ_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 (2010); 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 (2010). 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):

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

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

$$MRC = \frac{N}{N-k_N+2}\frac{1}{\psi_2^{k_N} k_N} \sum_{i=0}^{N-k_N+1} \bar{\boldsymbol{r}}_i \cdot \bar{\boldsymbol{r}}_i' - \frac{\psi_1^{k_N}}{\theta^2\psi_2^{k_N}}\hat{\Psi}, \tag{A.2}$$

where $\bar{\boldsymbol{r}}_i = (\bar{r}^{(1)}\ \bar{r}^{(2)}\ \cdots\ \bar{r}^{(d)})$ denotes the matrix of all pre-averaged return series and $\hat{\Psi}_N = \frac{1}{2N}\sum_{i=1}^{N} \boldsymbol{r}_{\tau_i}(\boldsymbol{r}_{\tau_i})'$. 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:

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

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.

Appendix B. CholCov asymptotics using pre-averaging estimators

The Modulated Realized Variance (MRV), defined in Appendix A, 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 **H** and **K** in Jacod et al. (2009), the estimates of g_{kk} conditional on no estimation error in previous estimates converge stably in law as*

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

where $\Gamma = \frac{4}{\psi_2^2}\left(\phi_{22}\theta\sigma^{*4} + 2\phi_{12}\frac{\sigma^{*2}\Psi^*}{\theta} + \phi_{11}\frac{\Psi^{*2}}{\theta^3}\right)$, Ψ^* 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 $\mathbb{E}\left(|\epsilon^j|^8\right) < \infty$ for all $j = 1, ..., d$. Define $\nabla = \hat{\Sigma}_{kk}^{-1}\left(1, -\hat{h}_{lk}\right)$, where $\hat{\Sigma}_{kk}$ is the $kk-th$ element of the MRC estimate. As $N \to \infty$ the distribution of h_{lk} conditional on all previous estimates is as follows*

$$N^{1/5}\left(\hat{h}_{lk} - h_{lk}\right) \xrightarrow{d} MN\left(\nabla'\left(\Psi_{lk}, \Psi_{kk}\right), \nabla\bar{\Gamma}\nabla'\right), \tag{B.1}$$

with

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

where

$$avar_{MRC} = \frac{2\Phi_{22}\theta}{\psi_2^2}\int_0^1 \Lambda(s)ds. \tag{B.3}$$

Λ is a $d \times d \times d \times d$ arrays with elements

$$\Lambda_s = \left\{\Sigma_{kk'}(s)\Sigma_{ll'}(s) + \Sigma_{kl'}(s)\Sigma_{lk'}(s)\right\}_{k,k'l,l,l'=1,...,d}.$$

Of course, avar* is as in Equation (B.3), but with $\Sigma = \Sigma^*$ and $\Psi = \Psi^*$.
Proof: see Theorem 6 in Christensen et al. (2010).

The estimate of the realized beta is inconsistent, due to a slight inconsistency in the MRC estimator when $\delta = 0.1$. The bias is very small in recent data and could be ignored. Alternatively, one could estimate the bias term to correct the MRC estimate, making it non-PSD, or estimate the realized beta bias term to correct it.
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 Σ^* and Ψ^*. 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.

Appendix C. 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	$\mathcal{D}$
Calender time grid (e.g. seconds)	t_i
Refresh-time grid on series $\measuredangle \subseteq \mathcal{D}$	$\tau_i^{\measuredangle}$
Number of observations refresh sample	$N^{\measuredangle}$
Number of refresh-times of constant correlation	$\varkappa$
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 variance	$D(s)$
Spot correlation matrix	$R(s)$
Lower diagonal matrix Cholesky	$H(s)$ with elements $h_{k,l}$
Diagonal matrix Cholesky	$G(s)$ with elements $g_{k,l}$
Matrices Palandri	$K(s)$
Elements in K	$\eta_{k,l}(s)$
Series	
Price observations	X_{τ_j}
Returns	$r_{\tau_j}^{(1)}, ..., r_{\tau_j}^{(d)}$
Cholesky projections	$f_{\tau_j}^{(1)}, ..., f_{\tau_j}^{(d)}$
Devolatized returns	$\hat{u}_{\tau_j}^{(1)}, ..., \hat{u}_{\tau_j}^{(d)}$
Devolatized and further standardized returns	$\tilde{u}_{\tau_j}^{(1)}, ..., \tilde{u}_{\tau_j}^{(d)}$