



HAL
open science

Positive semidefinite integrated covariance estimation, factorizations and asynchronicity

Kris Boudt, Sébastien Laurent, Asger Lunde, Rogier Quaadvlieg, Orimar Sauri

► To cite this version:

Kris Boudt, Sébastien Laurent, Asger Lunde, Rogier Quaadvlieg, Orimar Sauri. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 2017, 196 (2), pp.347-367. 10.1016/j.jeconom.2016.09.016 . hal-01505775

HAL Id: hal-01505775

<https://amu.hal.science/hal-01505775>

Submitted on 9 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity.

This Version: May 10, 2016

Kris Boudt^{a,b}, Sébastien Laurent^{c,d}, Asger Lunde^{e,f}, Rogier Quaadvlieg^{g,*}, Orimar Sauri^{e,f}

^a*Department of Business, Vrije Universiteit Brussel, Belgium.*

^b*Vrije Universiteit Amsterdam, Netherlands.*

^c*Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, France.*

^d*Aix-Marseille Graduate School of Management-IAE, France.*

^e*CREATES, Aarhus University, Denmark.*

^f*Department of Economics and Business, Aarhus University, Denmark.*

^g*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 of the covariance matrix in order to exploit the heterogeneity in trading intensities to estimate the different parameters sequentially with as many observations as possible. The estimator is positive semidefinite by construction. We derive asymptotic results and confirm their good finite sample properties by means of a Monte Carlo simulation. In the application we forecast portfolio Value-at-Risk and sector risk exposures for a portfolio of 52 stocks. We find that the dynamic models utilizing the proposed high-frequency estimator provide statistically and economically superior forecasts.

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

JEL: C10, C58

*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), osauri@econ.au.dk (Orimar Sauri)

1. Introduction

The availability of high-frequency data and a large variety of estimators harnessing their information, has enhanced the 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 estimators have sacrificed the latter (e.g. Hayashi and Yoshida, 2005; Zhang, 2011; Lunde et al., 2015).

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. Until recently, lower frequency returns, for instance sampled 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 that 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 a 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. (2015) 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 obtain a well-conditioned PSD matrix. However, due to the necessary transformations, we cannot sample over just two series for each el-

ement, but have to iteratively sample over an increasing set of series. Each bivariate estimate on the transformed series can be done with any estimator that is robust to the biases induced by ultra high-frequency data.

We derive the joint asymptotic properties of the realized covariance based CholCov estimator under the assumption of no microstructure noise and constant volatility. As the different elements of the CholCov are estimated using different time-grids, this is a non-trivial and new result. We find that the asymptotic covariance of the scaled CholCov parameter estimates is proportional to that of the classic realized covariance on synchronous returns and that the proportionality factor is a function of the overlap between the different grids used to estimate the CholCov parameters.

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 regularization technique to make the estimate positive semidefinite. Not only is it positive semidefinite, but simulations also show that the CholCov is on average better conditioned than its competitors, especially for larger dimensions. Finally, we conduct a simulation study which confirms that the asymptotic results also provide a good description of the finite sample distribution of the realized covariance based CholCov estimates in the absence of microstructure noise.

For the empirical application, we use our estimator to forecast portfolio Value-at-Risk and to estimate industry exposures for 52 assets. We forecast the covariance matrix using 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 derive its asymptotic properties. Section 5 presents a summary of 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.

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 $d \subseteq \mathcal{D}$. Each component process can be observed at different time points over the interval $[0, T]$. For simplicity we take $T = 1$ in this paper. For every $i = 1, \dots, d$, we denote by $T_i := \{0 = t_1^{(i)} < \dots < t_{N_i}^{(i)} \leq 1\}$ the observation times of the i -th element in X . X is driven by the efficient log-price Y , a Brownian semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq 1}, \mathbb{P})$ satisfying the usual conditions of right-continuity and completeness:

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

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:

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

where $\epsilon_t^{(i)}$ is microstructure noise and $Y_t^{(i)}$ is the i -th component of Y . In this paper, $\epsilon_t = (\epsilon_t^{(1)}, \dots, \epsilon_t^{(d)})'$ is assumed to be covariance stationary and have zero mean. We achieve noise robustness by means of pre-averaging and refer the reader to Christensen et al. (2010) for the precise assumptions on the serial correlation of the noise and its cross-correlation with the latent price.

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

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

which under (1) coincides with the quadratic covariation of Y , i.e. $[Y](1) = \text{plim}_{n \rightarrow \infty} \sum_{j=1}^n (Y_{j/n} - Y_{(j-1)/n})(Y_{j/n} - Y_{(j-1)/n})' = \text{ICov}$.

Estimation of the off-diagonal elements of the ICov typically requires synchronization of the data.¹ 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 off-diagonal elements of the ICov separately using synchronized pairs of price processes, but the result is not ensured to be positive semidefinite (e.g. Lunde et al., 2015). We propose a third approach that uses a factorization of the spot covariance 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 require that the parameters defining the spot covariance 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 amongst others the Cholesky decomposition, which we will focus on.

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. Palandri (2009) uses an iterative Cholesky to reduce the dimension of multivariate GARCH models. The use of the Cholesky decomposition is the only similarity with our work. Their methods and goals are different from ours. We construct an ex-post covariance estimator, not a time-series parametric volatility model.

By means of the Cholesky decomposition, the spot covariance matrix can be uniquely split into

$$\Sigma(s) = H(s)G(s)H(s)', \quad (4)$$

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

¹There are methods available that circumvent this necessity, such as the estimator of Hayashi and Yoshida (2005) and its noise-robust version proposed by Koike (2013), or the missing value approach of Shephard and Xiu (2013) and Corsi et al. (2015).

ically,

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

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

$$\Sigma = HGH' = \begin{bmatrix} g_{11} & h_{21}g_{11} & h_{31}g_{11} \\ h_{21}g_{11} & h_{21}^2g_{11} + g_{22} & h_{21}h_{31}g_{11} + h_{32}g_{22} \\ h_{31}g_{11} & h_{21}h_{31}g_{11} + h_{32}g_{22} & h_{31}^2g_{11} + h_{32}^2g_{22} + g_{33} \end{bmatrix}. \quad (5)$$

It follows from this that the H and G matrices are linked to the elements of Σ as follows:

$$g_{kk} = \Sigma_{kk} - \sum_{m=1}^{k-1} h_{km}^2 g_{mm} \quad (6)$$

$$h_{kl} = \frac{1}{g_{ll}} \left(\Sigma_{kl} - \sum_{m=1}^{k-1} h_{km} h_{lm} g_{mm} \right), \quad (7)$$

for $k > 1$ with $g_{11} = \Sigma_{11}$. The (k, l) -th element of the covariance 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. The estimator

In order to apply the Cholesky decomposition in the estimation of ICov, we first need to define how the returns are computed and, since the procedure will be sequential, how the covariance matrix is ordered. Therefore this section is divided into four parts. The concept of refresh-time sampling is reviewed in Subsection 3.1. The liquidity criterion used to structure the covariance matrix is introduced in Subsection 3.2. Finally, the CholCov estimator and some practically relevant extensions are presented respectively in Subsections 3.3 and 3.4.

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. Refresh-times for $t \in [0, 1]$ and the set of assets $d \subseteq \mathcal{D}$ are defined as follows. The first refresh-time occurs at $\tau_1^d = \max(t_1^{(i)}, \forall i \in d)$, and subsequent refresh-times at $\tau_{j+1}^d = \max(t_m^{(i)} | t_m^{(i)} > \tau_j^d, m = 1, \dots, N_i, \forall i \in d)$. The resulting refresh-time grid is of course a function of the series over which they are sampled. In our setting, stocks are ranked on liquidity and of particular interest will be the refresh-time grid obtained for the first k assets, i.e. $d = \{1, \dots, k\}$. Denote the set of corresponding refresh-times as $\mathcal{T}_k := \{0 = \tau_0^k < \tau_1^k < \dots < \tau_{\mathcal{N}^k}^k = 1\}$. We denote the returns of grid \mathcal{T}_k by $r_j(\mathcal{T}_k) = X_{\tau_j^k} - X_{\tau_{j-1}^k}$, with $j = 1, \dots, \mathcal{N}^k$. Finally, denote the durations as $\Delta_j(\mathcal{T}_k) = \tau_j^k - \tau_{j-1}^k$. 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, $\mathcal{N}^k \leq \min_{i \in d} N_i$, 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 largely determined by the least liquid - in the sense of frequent trades - stocks. Including a single illiquid stock 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. (2015) 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.

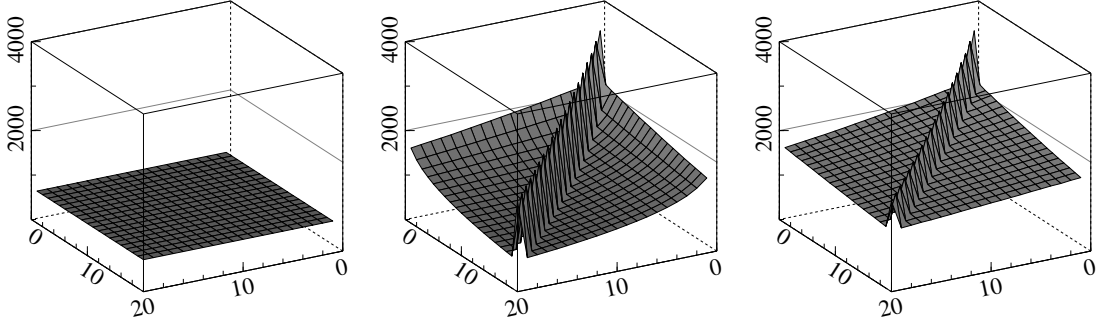
Here, we use the Cholesky decomposition to exploit the heterogeneity in trading intensities and estimate the different parameters sequentially, using as much data as possible. The decomposition holds for any PSD matrix, which is by definition true for the covariance matrix, and the recombined results 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 obtained for each element using three different sampling methods. We consider 20 assets, where each asset is slightly less liquid than the last, with trading activity of asset i governed by a Poisson process with intensity $\lambda_i = 10 + i/4$. The leftmost picture depicts a situation in which refresh-time sampling is applied to all assets at once. The rightmost picture depicts pairwise sampling, which is done for the Composite estimators of Lunde et al. (2015). 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 and therefore the data-loss is 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. Specifically, for element (k, l) we synchronize series $\{1, \dots, l, k\}$. 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 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_i . However, there are many scenarios where this would not lead to a large number of observations when sampling over many series. Instead we

Figure 1: Number of observations available for each element in estimation of Σ , 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 20×20 covariance matrix for different sampling schemes. The observation frequencies follow independent Poisson processes with parameter λ ranging uniformly from 10 and 15. Stocks are ranked based on λ . For more details, see Section 5. From left to right: Sampling over all series, sequential sampling and pairwise sampling.

use the squared duration as a liquidity criterion²

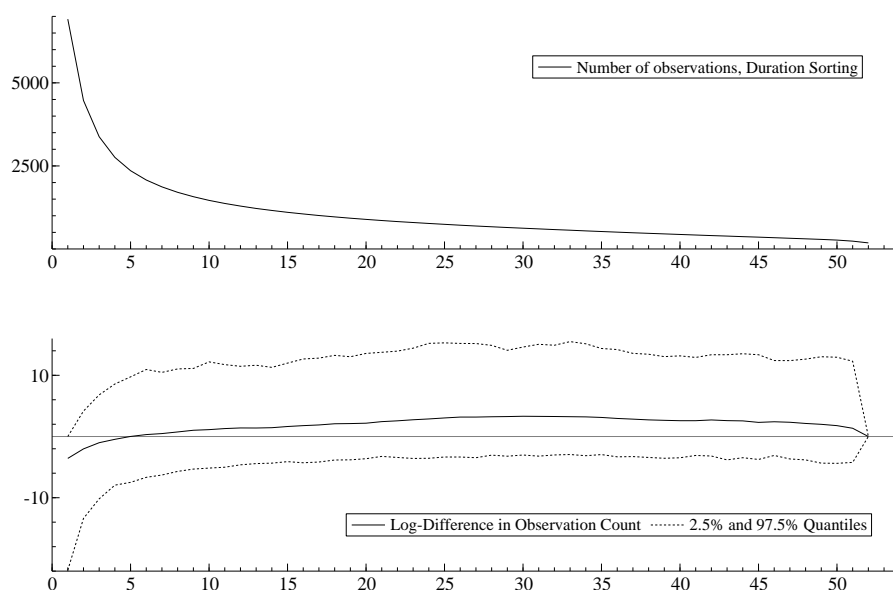
$$\sum_{j=0}^{N_i} \left(\Delta_j^{(i)} \right)^2, \quad (8)$$

where $\Delta_j^{(i)}$ is each individual series' duration. This criterium is not just based on the number of intraday returns available, but also on the degree to which the returns are spread out over the day. We call our method duration sorting, as opposed to observation count sorting proposed in Hautsch et al. (2012).

Different sorting rules will lead to different sets being synchronized at any iteration in the sequential estimation. To illustrate the effect of duration sorting compared to observation count sorting, we preview results of our empirical application where we synchronize 52 series on 1499 days in Figure 2. Each day, the series are sorted according to the two methods. Next, for each of the methods, an increasing set of series is synchronized, from a single series, up to the full 52. As such, the x-axes in Figure 2 denote the number of series that is synchronized. For each set size, we therefore have a distribution over the different days. We plot the average amount of observations per set size for duration sorting in the top plot. In the bottom plot, we display the mean and the 2.5 and 97.5%

²In a previous version we have proposed relative duration $\sum_{j=0}^{N_i} \left(\Delta_j^{(i)} - \frac{1}{\max_i N_i} \right)^2$, which leads to sorts identical those based on Equation (8) in almost all cases. We have also considered Absolute Duration, which leads to inferior sorting.

Figure 2: Duration sorting and observation count sorting, as a function of the number of series synchronized.



Note: The top graph shows the number of observations available after synchronization over an increasing number of series from 1 to 52. The bottom graph shows the percentage gains in number of observations from using duration sorting compared to observation count sorting.

empirical quantiles of the log-difference of the number of observations between the two sampling techniques. As such, in the bottom graph the y-axis can be interpreted as percentage increases in number of observations by using duration sorting compared to observation count sorting.

The figure shows that duration sorting leads to a smoothing of available observations across the set of series. When synchronizing just the first one to six series, there are on average fewer observations with duration sorting. However, starting from the synchronization of the first seven series, up until the full set, the average number of observations using duration sorting is higher. This has large advantages. As the top graph shows, for small sets, there are plenty of observations available, while when more series are synchronized, observations are a lot less frequent. The number of observations using duration sorting is on average higher by 2 to 3%. Additionally, the distribution is skewed towards higher gains than losses, with the duration 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. Finally, once all 52 series are synchronized there is no difference between the sorting rules, as the sets are the same for both.

3.3. The CholCov

This section presents our estimator of the Integrated Covariance, which achieves positive semidefiniteness by means of the Cholesky decomposition. Instead of directly estimating the elements in Σ , a sequential regression algorithm is proposed to estimate the elements of the G and H matrices, which are then combined into a final estimate of Σ , denoted as the CholCov:

$$\text{CholCov} = \hat{H}\hat{G}\hat{H}', \quad (9)$$

with \hat{G} and \hat{H} the corresponding estimates for G and H . The computational cost of the algorithm is compensated by the expected gains in efficiency in the presence of asynchronicity.

The inspiration for the algorithm comes from rewriting the returns in a block of locally constant volatility, $\Sigma^0 = H^0 G^0 H^{0'}$, as a factor regression model. For simplicity, we assume the block corresponds to the $[0,1]$ interval.³ Denote the vector of latent noise-free returns by $\tilde{r}_j(T) = Y_{t_j} - Y_{t_{j-1}}$, with $T = \{0 = t_1 < \dots < t_N = 1\}$, 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 $\Delta_j(T)\Sigma^0$. We can further define, omitting the grid-dependence, $\tilde{f}_j = (H^0)^{-1}\tilde{r}_j \sim N(0, \Delta_j G^0)$. Since H^0 is triangular, we obtain that each component in the vector of returns $\tilde{r}_j = H^0 \tilde{f}_j$, is an explicit function of the components with a lower index. More precisely, $\tilde{f}_j^{(1)}(T) = \tilde{r}_j^{(1)} \sim N(0, \Delta_j g_{11}^0)$, and for $k = 2, \dots, d$,

$$\begin{aligned} \tilde{r}_j^{(k)} &= h_{k1}^0 \tilde{f}_j^{(1)} + \dots + h_{k(k-1)}^0 \tilde{f}_j^{(k-1)} + \tilde{f}_j^{(k)} \\ \tilde{f}_j^{(k)} &\sim N(0, \Delta_j g_{kk}^0). \end{aligned} \quad (10)$$

The g_{kk}^0 and h_{kl}^0 elements are therefore simply the residual variances of the factors and the beta coefficients of a regression of $\tilde{r}_j^{(k)}$ on $\tilde{f}_j^{(l)}$:

$$g_{kk}^0 = [\tilde{f}^{(k)}] \quad (11)$$

$$h_{kl}^0 = [\tilde{r}^{(k)}, \tilde{f}^{(l)}] / [\tilde{f}^{(l)}], \quad (12)$$

for $l < k$, where the $[\cdot, \cdot]$ notation denotes the quadratic (co)variation of its element(s).

³The one day block simplifies the exposition and is also the one used in the implementation. Of course, the $[0,1]$ interval can be subdivided in infinitely many blocks and the resulting CholCov estimate would be the sum over all these local CholCov estimates.

This suggests then to estimate the CholCov parameters iteratively on an expanding set of synchronized price series. Let \mathcal{T}_k be the set of refresh-times for the first k -assets and assume that $\mathcal{T}_k \subseteq \mathcal{T}_j$ for $j = 1, \dots, k$.

CholCov estimation algorithm.

1. First sort the series in terms of decreasing liquidity according to (8), such that series 1 is the most liquid, and series d the least.

2. For $k = 1, \dots, d$:

Apply refresh-time on $\kappa = \{1, \dots, k\}$ to obtain the grid \mathcal{T}_k .

Set $\hat{f}_j^{(1)}(\mathcal{T}_k) = r_j^{(1)}(\mathcal{T}_k)$.

For $u = 1, \dots, k$ and $v = 1, \dots, u - 1$, compute:

$$\bar{g}_{uu}(\mathcal{T}_k) = \sum_{j=1}^{\mathcal{N}^k} [\hat{f}_j^{(u)}(\mathcal{T}_k)]^2, \quad (13)$$

$$\bar{h}_{uv}(\mathcal{T}_k) = \frac{\sum_{j=1}^{\mathcal{N}^k} r_j^{(u)}(\mathcal{T}_k) \hat{f}_j^{(v)}(\mathcal{T}_k)}{\sum_{j=1}^{\mathcal{N}^k} [\hat{f}_j^{(v)}(\mathcal{T}_k)]^2}, \quad (14)$$

where

$$\hat{f}_j^{(v)}(\mathcal{T}_k) := r_j^{(v)}(\mathcal{T}_k) - \sum_{w=1}^{v-1} \bar{h}_{vw}(\mathcal{T}_k) \hat{f}_j^{(w)}(\mathcal{T}_k). \quad (15)$$

3. Set the diagonal elements of \hat{G} and lower diagonal elements of \hat{H} as

$$\hat{g}_{kk} = \bar{g}_{kk}(\mathcal{T}_k) \quad (16)$$

$$\hat{h}_{kl} = \bar{h}_{kl}(\mathcal{T}_k), \quad (17)$$

for $k = 1, \dots, d$ and $l = 1, \dots, k - 1$, and compute the (realized covariance based) CholCov as $\hat{H}\hat{G}\hat{H}'$.

3.4. Using CholCov in practise

The CholCov estimator given by (16)-(17) uses realized (co)variance estimates for the g_{kk} and h_{kl} parameters. This choice of estimator is no longer adequate in the presence of microstructure noise, where we only observe a noisy counterpart of $\tilde{r}_j^{(k)}$ and $\tilde{f}_j^{(k)}$, namely the refresh return

$$r_j^{(k)} = \tilde{r}_j^{(k)} + \epsilon_j^{(k)} - \epsilon_{j-1}^{(k)}, \quad (18)$$

and the corresponding noise-contaminated factor

$$f_j^{(l)} = \tilde{f}_j^{(l)} + (\epsilon_j^{(l)} - \epsilon_{j-1}^{(l)}) - \sum_{a=1}^{l-1} h_{la}(\epsilon_j^{(a)} - \epsilon_{j-1}^{(a)}), \quad (19)$$

for $l > 1$. The CholCov estimates of g_{kk} and h_{kl} that we recommend to use in practice therefore result from estimating the quadratic (co)variation in (11)-(12) with noise-robust integrated variance (IV) and beta estimators implemented on synchronized price series.⁴ In the simulation study and empirical application we will use the Modulated Realized Covariance (MRC) pre-averaging estimator of Christensen et al. (2010) because of its robustness to microstructure noise, its ease of implementation and its flexibility. We denote the corresponding estimate as the *mrcCholCov* or *CholCov* when there is no confusion possible.

Other approaches that are robust to microstructure noise include the two time scale estimator of Zhang (2011), the kernel estimator in Barndorff-Nielsen et al. (2011), the likelihood based methods in Aït-Sahalia et al. (2010), the local spectral estimator in Bibinger and Reiß (2014), the local method of moments estimator in Bibinger et al. (2014) as well as advances by Peluso et al. (2014) and Shephard and Xiu (2013), among others.

For the estimation of h_{kl} , the CholCov estimator proposed above requires synchronizing the prices of assets $1, \dots, k$. When the factors are orthogonal, a consistent estimator for h_{kl} could also readily be obtained by synchronizing only the series $1, \dots, l$ needed to compute the factor f^l , together with the observed prices for asset k . In finite samples, the factors are no longer orthogonal when using these grids on a smaller number of series. Our simulation analysis shows that substantial efficiency gains are obtained by estimating h_{kl} on the refresh-time grid for $\{1, \dots, l, k\}$ versus the estimation using the grid obtained for $\{1, \dots, k\}$. The corresponding CholCov* estimation algorithm is then as follows:

CholCov* Estimation algorithm.

1. First sort the series in terms of decreasing liquidity according to (8), such that series 1 is the most liquid, and series d the least.

2. *Step 1:*

$$\text{Set } f_j^{(1)}(T^1) = r_j^{(1)}(T^1).$$

⁴The realized beta is computed by estimating a 2×2 realized covariance matrix $\hat{\Sigma}$. The realized beta is then defined as $\hat{h}_{kl} = \hat{\Sigma}_{kl} / \hat{\Sigma}_{ll}$ (Andersen et al., 2006).

Estimate \hat{g}_{11} using an IV estimator on $f_j^{(1)}(T^1)$.

3. *Step 2:*

Apply refresh-time on $b = \{1, 2\}$ and denote the corresponding grid by \mathcal{T}_b .

Estimate \hat{h}_{21} as the realized beta of $r_j^{(2)}(\mathcal{T}_b)$ on $f_j^{(1)}(\mathcal{T}_b)$.

Estimate \hat{g}_{22} using an IV estimator on $f_j^{(2)}(\mathcal{T}_b) = r_j^{(2)}(\mathcal{T}_b) - \hat{h}_{21}f_j^{(1)}(\mathcal{T}_b)$.

4. *Step 3:*

Apply refresh-time on $c = \{1, 3\}$ to obtain the grid \mathcal{T}_c .

Estimate \hat{h}_{31} as the realized beta of $r_j^{(3)}(\mathcal{T}_c)$ on $f_j^{(1)}(\mathcal{T}_c)$.

Apply refresh-time on $d = \{1, 2, 3\}$ to obtain the grid \mathcal{T}_d .

Compute $f_j^{(2)}(\mathcal{T}_d) = r_j^{(2)}(\mathcal{T}_d) - \hat{h}_{21}f_j^{(1)}(\mathcal{T}_d)$.

Estimate \hat{h}_{32} as the realized beta of $r_j^{(3)}(\mathcal{T}_d)$ on $f_j^{(2)}(\mathcal{T}_d)$.

Estimate \hat{g}_{33} using an IV estimator on $f_j^{(3)}(\mathcal{T}_d) = r_j^{(3)}(\mathcal{T}_d) - \hat{h}_{32}f_j^{(2)}(\mathcal{T}_d) - \hat{h}_{31}f_j^{(1)}(\mathcal{T}_d)$.

5. *Step $k=4$ to d :*

For $l = 1, \dots, (k-1)$:

Create the grid τ^z by applying refresh-time on $z = \{1, \dots, l, k\}$ to obtain the grid \mathcal{T}_z .

Extract the factors $f_j^{(m)}(\mathcal{T}_z) = r_j^{(m)}(\mathcal{T}_z) - \sum_{n=1}^{m-1} \hat{h}_{mn}f_j^{(n)}(\mathcal{T}_z)$, for $m = 1, \dots, l$ at the new grid.

Estimate \hat{h}_{kl} as the realized beta of $r_j^{(k)}(\mathcal{T}_z)$ on $f_j^{(l)}(\mathcal{T}_z)$.

Estimate \hat{g}_{kk} using an IV estimator on $f_j^{(k)}(\mathcal{T}_z) = r_j^{(k)}(\mathcal{T}_z) - \sum_{n=1}^{k-1} \hat{h}_{kn}f_j^{(n)}(\mathcal{T}_z)$.

Finally, we also recommend to use a strip and replace method such that the diagonal elements of the CholCov are estimated using all available observations. More precisely, denote by \hat{D} the diagonal matrix with the ratio between the integrated variance estimates and the corresponding diagonal elements of $\hat{H}\hat{G}\hat{H}'$. Then the strip and replace version of the CholCov is

$$CholCov^\bullet = \hat{D}^{1/2} \hat{H} \hat{G} \hat{H}' \hat{D}^{1/2}. \quad (20)$$

In the simulations and application we use the MRC pre-averaging estimator of Christensen et al. (2010) for the estimation of g_{kk} and h_{kl} and denote this by $mrcCholCov^\bullet$ or $CholCov$ when there is no confusion possible.

4. Asymptotic results

In this section, we derive the asymptotic distribution of the *CholCov* estimator in (16)-(17). Throughout this section, we assume that the refresh-times \mathcal{T}_k satisfy $\mathcal{T}_k \subseteq \mathcal{T}_i$, for $i = 1, \dots, k$, and $\mathcal{T}_{k+1} \subseteq \mathcal{T}_k$, such that the returns on the refresh-time grid \mathcal{T}_{k+1} can be written as a linear combination of the returns on the grid \mathcal{T}_k . We further simplify the model by assuming that the underlying diffusion process has constant volatility, no drift and that the prices are observed without microstructure noise, i.e.

$$dY_t = \sigma dW_t. \quad (21)$$

We show that even in this case, obtaining the asymptotic distribution of the parameters estimated sequentially on different time grids is non-trivial.

The main idea of our proof consists in showing that the k th row of the *CholCov* estimator is asymptotically equivalent to the k th row of the realized covariance of Y computed for the mesh \mathcal{T}_k . It follows that the *CholCov* estimated on non-synchronous series is consistent and marginally normal. Then, using the results in Nualart and Peccati (2005) and Peccati and Tudor (2005), we derive the joint distribution.

In order to obtain this result, we proceed in four steps:

1. First, we show that in the synchronized case, the pair $(\widehat{H}, \widehat{G})$ given by (16)-(17) is the Cholesky decomposition of the MLE of $\Sigma^0 := \sigma\sigma'$. This implies that in the synchronized case, *CholCov* is just the MLE estimator of Σ^0 . This result is used in Proposition 1, where consistency and asymptotic normality of the pair $(\widehat{H}, \widehat{G})$ is established.
2. Second, for non-synchronous data with $\mathcal{T}_{k+1} \subseteq \mathcal{T}_k$, we note that the k th row of *CholCov* corresponds to the k th row of the MLE of Σ^0 based on $\{Y_t : t \in \mathcal{T}_k\}$.
3. Third, let $(A)_k$ denote the k th row of a matrix A , and *RCov* the realized covariance of the process Y in the mesh \mathcal{T}_k . We introduce Assumption 1 to guarantee that for $k = 1, \dots, d$

$$(\text{CholCov})_k = (\text{RCov})_k + o_{\mathbb{P}}\left(\mathcal{N}_k^{1/2}\right).$$

Since the mapping $(H, G) \mapsto HGH'$ is bijective and continuously differentiable, it follows that under Assumption 1, $(\text{CholCov})_k$ is consistent and asymptotically normal. These results are presented in Proposition 2.

4. Finally, the joint asymptotic distribution of *CholCov* (based on the limit distribution of *RCov*) is derived in Theorem 1 using the results of Nualart and Peccati (2005) and Peccati and Tudor (2005).

All proofs are provided in Appendix B.

4.1. Preliminaries

In this section, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq 1}, \mathbb{P})$ denotes a filtered probability space satisfying the usual conditions of right-continuity and completeness and where the observed prices are generated by the Brownian martingale without drift and constant volatility in (21). Furthermore, we assume that $\mathcal{F} = \sigma((W_t)_{0 \leq t \leq 1})$, the σ -algebra generated by our underlying Brownian motion.

Given two sequences $\{a_n\}_{n \in \mathbb{N}}, \{b_n\}_{n \in \mathbb{N}}$, we write $a_n = O(b_n)$ if and only if there exist $M > 0$ and $n_0 \in \mathbb{N}$, such that for any $n \geq n_0$, $|a_n| < M|b_n|$ for any $n \in \mathbb{N}$. When $b_n \neq 0$ for any $n \in \mathbb{N}$, this is equivalent to $\limsup_{n \rightarrow \infty} a_n/b_n < \infty$. Furthermore, we write $a_n = o(b_n)$ if and only if for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that for any $n \geq n_0$, $|a_n| < \varepsilon|b_n|$. Equivalently, $a_n = o(b_n)$ if and only if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$, whenever $b_n \neq 0$. The notation $a_n = O_{\mathbb{P}}(b_n)$ and $a_n = o_{\mathbb{P}}(b_n)$ is defined in a similar way by replacing usual convergence with convergence in probability.

Let \mathbb{M}_d^+ be the space of symmetric positive definite matrices with real entries. We embed \mathbb{M}_d^+ with the Frobenius norm such that if $A \in \mathbb{M}_d^+$, then $\|A\| = \text{tr}(A^T A)^{1/2}$. The Cholesky decomposition is unique on \mathbb{M}_d^+ . This means that the mapping $\Sigma(H, G) = HGH'$, creates a bijection between the convex space

$$\mathbb{M}_{d, \text{chol}}^+ = \{(H, G) : G \text{ diagonal with } g_{kk} \geq 0, \text{ and } H \text{ lower diagonal with ones in its diagonal}\},$$

and \mathbb{M}_d^+ . Let I_k be the $k \times k$ identity matrix and Π_k the projection matrix, which, for a d -dimensional vector $x = (x_1, \dots, x_d)'$ on $x(k) := (x_1, \dots, x_k)'$. Furthermore, if A is $d \times d$ matrix and $1 \leq k \leq d$, $A(k)$ will denote the projection of A into its first submatrix of dimension $k \times k$, i.e. $A(k) = (A_{i,j})_{i,j=1, \dots, k}$. In particular, if $A \in \mathbb{M}_d^+$,

$$A(k) = \Pi_k A \Pi_k'. \quad (22)$$

We use a row version of the vech operator that stacks the rows of the lower-triangular part of a

d -dimensional matrix A into a vector, as follows:

$$\text{rvech}(A) = (a_{11}, a_{21}, a_{22}, \dots, a_{k1}, a_{k2}, \dots, a_{kk-1}, a_{kk}, \dots, a_{dd-1}, a_{dd})'.$$

Note that

$$\text{rvech}(A(k)) = L(\Pi_k \otimes \Pi_k)MB^{-1}\text{rvech}(A), \quad (23)$$

with L and M such that $\text{vech}(A) = L\text{vec}(A)$ and $\text{vec}(A) = M\text{vech}(A)$ and B satisfying $\text{rvech}(A) = B\text{vech}(A)$. Obviously rvech is an injective continuously differentiable linear mapping from \mathbb{M}_d^+ into a convex subset of $\mathbb{R}^{d(d+1)/2}$.

Throughout the following, we use versions of $[(H - I_d) + G]$, where all elements on a single row are estimated on the same grid. The parameters on a lower row are estimated using refresh-times of a larger number of assets than those belonging to an upper row. The rvech of this matrix therefore stacks the elements in such a way that elements on the same grid are adjacent.

Finally, for simplicity of exposition, for the rest of this section we assume that the observation times are deterministic. For the general case, see Remark 4 below.

4.2. The case of synchronous observations

Here we assume that the observation times are synchronous, i.e $T := \mathcal{T}_1 = \mathcal{T}_2 = \dots = \mathcal{T}_d$. We consider first the case of n equispaced returns generated by (21). In this case, the log-likelihood function is:

$$\bar{Q}_n(\Sigma^0) := -\frac{1}{2} \{ \ln(|\Sigma^0|) + \text{tr}[(\Sigma^0)^{-1}\bar{\Sigma}_n] \}, \quad n \in \mathbb{N}, \quad (24)$$

where $\bar{\Sigma}_n$ is the realized covariance estimator using all available prices, i.e,

$$\bar{\Sigma}_n = \bar{\Sigma} := \sum_{j=1}^n (Y_{t_j^n} - Y_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n})'. \quad (25)$$

It is well known that $\bar{\Sigma}_n$ maximizes \bar{Q}_n on \mathbb{M}_d^+ and, under \mathbb{P} , is strongly consistent for Σ^0 , that is $\bar{\Sigma}_n \xrightarrow{a.s.} \Sigma^0$, as $n \rightarrow \infty$. On the other hand, since $\bar{\Sigma}_n \in \mathbb{M}_d^+$ almost surely, there exists a unique pair of matrices, \bar{G}_n and \bar{H}_n , diagonal and lower diagonal with ones on its diagonal, respectively, such that

$$\bar{\Sigma}_n = \bar{H}_n \bar{G}_n \bar{H}_n', \quad \text{a.s.}$$

It follows from this that, $\bar{g}_{11} = \bar{\Sigma}_{11}$, $\bar{h}_{i1} = \frac{\bar{\Sigma}_{i1}}{\bar{g}_{11}}$ for $i = 2, \dots, d$, and for $k \geq 2$ and $k > l$

$$\bar{g}_{kk} = \bar{\Sigma}_{kk} - \sum_{m=1}^{k-1} \bar{h}_{km}^2 \bar{g}_{mm}, \quad (26)$$

$$\bar{h}_{kl} = \frac{1}{\bar{g}_{ll}} \left(\bar{\Sigma}_{kl} - \sum_{m=1}^{l-1} \bar{h}_{kl} \bar{h}_{ml} \bar{g}_{mm} \right). \quad (27)$$

Proposition 1 *Suppose that $T_1 = T_2 = \dots = T_d$ and $\Delta_i(T) = 1/n$. Let $(\hat{H}, \hat{G}) \in \mathbb{M}_{d, \text{chol}}^+$ be the matrices associated to (16) and (17). Then (\hat{H}, \hat{G}) is the Cholesky decomposition of $\bar{\Sigma}$ (the MLE estimator of Σ^0). In particular, (\hat{H}, \hat{G}) is strongly consistent for (H^0, G^0) under \mathbb{P} and*

$$n^{1/2} (\hat{q} - q^0) \xrightarrow{d} N(0, \Phi), \quad \text{as } n \rightarrow \infty,$$

where $\hat{q} := \text{rvech} \left[(\hat{H} - I_d) + \hat{G} \right]$, $q^0 := \text{rvech} [(H^0 - I_d) + G^0]$ and $\Phi \in \mathbb{M}_{d(d+1)/2}^+$ is the inverse of the Fisher's information associated to \bar{Q}_n as a function of q^0 .

General formulas for the score function, the Hessian matrix, and the Fisher information for the MLE of q^0 are given in Appendix C. In particular, we show that the estimation error of the MLE for the g_{kk}^0 parameters is asymptotically uncorrelated to that of the MLE for the h_{kl}^0 parameters, and that the estimation errors of h_{kl}^0 and h_{mn}^0 are only correlated if $l = n$. As such, the Fisher matrix and its inverse are block diagonal. An unsurprising result is that the asymptotic variance of the MLE for the g_{kk}^0 elements, which are computed as realized variances, is equal to $2(g_{kk}^0)^2$.

Remark 1 Observe that in the previous proposition (\hat{H}, \hat{G}) is the Cholesky decomposition of the realized covariance associated to the mesh T . This property only holds when $\Delta_i(T) = 1/n$. For the general case (non-equally spaced mesh), if $T = \{0 = t_0^n < t_1^n < \dots < t_n^n = 1\}$, then the MLE (\tilde{H}, \tilde{G}) is the Cholesky decomposition of

$$\tilde{\Sigma} = \sum_{j=1}^n \frac{1}{n \Delta_j(T)} (Y_{t_j^n} - Y_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n})', \quad (28)$$

where $\Delta_j(T) = t_j^n - t_{j-1}^n$, which of course does not coincide with the realized covariance. However, we can resolve this situation by imposing asymptotic behavior on the observation times. Indeed, let $\hat{\Sigma}$ be the realized covariance associated to the mesh T and $\tilde{\Sigma}$ the MLE estimator of (28). Trivially,

we have

$$\left\| \widehat{\Sigma} - \widetilde{\Sigma} \right\| \leq \sup_j \left| \frac{1}{\Delta_j(T)n} - 1 \right| \left\| \widehat{\Sigma} \right\|. \quad (29)$$

Therefore, if

$$\sup_j |\Delta_j(T)n - 1| = o\left(n^{-1/2}\right), \quad \text{as } n \rightarrow \infty, \quad (30)$$

we have that $\widehat{\Sigma} = \widetilde{\Sigma} + o_{\mathbb{P}}(n^{1/2})$. Therefore, if (30) holds, then the conclusions in Proposition 1 also hold for non-equispaced data. Finally, let us remark that (30) implies $|T| := \sup_j (t_j^n - t_{j-1}^n) \sim 1/n$ as $n \rightarrow \infty$. This means that the convergence of $\widehat{\Sigma}$ to Σ^0 is not only in probability but also almost surely. See De la Vega (1974) for more details.

4.3. The case of non-synchronous observations

Recall that we are assuming the observation times to be deterministic. In order to study the non-synchronized case, in analogy to Remark 1, we introduce the following additional assumption:

Assumption 1 For every $k = 1, \dots, d-1$, we assume $T_{k+1} \subseteq T_k$ (or $\mathcal{T}_{k+1} \subseteq \mathcal{T}_k$). Further, for any $k, i = 1, \dots, d$, as $n \rightarrow \infty$, $N_k \rightarrow \infty$, $N_k/N_i = K_{k,i} + o(1)$ for some $K_{k,i} > 0$, and

$$\sup_j |\Delta_j^k(T)N_k - 1| = o\left(N_k^{-1/2}\right), \quad \text{as } n \rightarrow \infty. \quad (31)$$

Remark 2 Assumption 1 is standard in the non-synchronous framework, see e.g. Mykland (2010), Hayashi and Yoshida (2008) and Christensen et al. (2013). Furthermore, as pointed out in Remark 1, our assumption on the speed of convergence on (31) is to control the asymptotic error between the CholCov estimator and the quadratic covariation of Y , which arises from the choice of a maximum likelihood based estimator.

In the presence of non-synchronous data, the estimation is done iteratively on expanding subsets. For a given refresh-time grid \mathcal{T}_k on which we estimate the Cholesky parameters G_k^0 and H_k^0 of the projected covariance matrix $\Sigma^0(k) = \Pi_k \Sigma^0 \Pi_k'$ using (13) and (14), the observations are synchronous and the following result follows directly from the previous results.

Proposition 2 Let Assumption 1 hold. Let also $(\hat{g}_{kk})_{k=1, \dots, d}$ and $(\hat{h}_{kl})_{k=1, \dots, d, l < k}$ be defined as in (16) and (17). Let us denote

$$\hat{q}_k := A_k \text{rvech} \left[\left(\widehat{H}(k) - I_k \right) + \widehat{G}(k) \right], \quad q_k^0 := A_k \text{rvech} \left[\left(H^0(k) - I_k \right) + G^0(k) \right],$$

with $A_k = \begin{bmatrix} 0 & I_k \end{bmatrix}$, i.e. the k -th row of $(\widehat{H}(k) - I_k) + \widehat{G}(k)$ and $(H^0(k) - I_k) + G^0(k)$, respectively. Then, for every $k = 1, \dots, d$, $\hat{q}_k \xrightarrow{\mathbb{P}} q_k^0$ and

$$\sqrt{N_k} [\hat{q}_k - q_k^0] \xrightarrow{d} N(0, \bar{\Phi}_k), \quad \text{as } n \rightarrow \infty,$$

where $\bar{\Phi}_k = A_k \Phi_k A_k'$, with $\Phi_k = \Pi_{k(k+1)/2} \Phi \Pi'_{k(k+1)/2}$ and Φ defined as in Proposition 1.

4.4. A CLT for the CholCov

In this subsection, within the framework of Assumption 1, we derive the asymptotic distribution of our *CholCov* estimator. In Appendix A, we provide a summary of the theory used to proof our result.

Under the notation of Proposition 2, we have that $q_k = A_k q$. Furthermore, as the mapping $(H, G) \mapsto \text{rvech}(HGH^T)$ is a continuously differentiable bijection between $\mathbb{M}_{d, \text{chol}}^+$ and $\mathbb{R}^{d(d+1)/2}$, by the delta method, q is asymptotically normal if and only if the random vector $\left(A_k \text{rvech} \left[\widehat{\Sigma}(k) \right] \right)_{k=1, \dots, d}$ is asymptotically normal as well. Moreover, under Assumption 1, $\text{rvech} \left[\widehat{\Sigma}(k) \right] = \text{rvech} \left[\widetilde{\Sigma}(k) \right] + o_{\mathbb{P}} \left(N_k^{-1/2} \right)$, with $\widehat{\Sigma}(k)$ the realized covariance of the process $\pi_k(Y)$ for the mesh \mathcal{T}_k . Hence, the asymptotic properties of q are the ones of $\left(A_k \text{rvech} \left[\widehat{\Sigma}(k) \right] \right)_{k=1, \dots, d}$, which we derive in the following theorem.

Theorem 1 *Let Assumption 1 hold. For every $k = 1, \dots, d$, let $\beta_k = A_k \text{rvech} \left[\sqrt{N_k} \left(\widehat{\Sigma}(k) - \Sigma^0(k) \right) \right]$ be the k -th row of $\sqrt{N_k} \left(\widehat{\Sigma}(k) - \Sigma^0(k) \right)$. Then as $n \rightarrow \infty$, the vector $\beta = (\beta_k)_{k=1, \dots, d}$ converges in distribution to a normal random vector with covariance structure given by*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\beta_k(i) \beta_{k+l}(j)] := \Psi((k, i), (k+l, j)) \rho_{k, k+l}, \quad (32)$$

where, for every $k = 1, \dots, d$, $l = 1, \dots, d - k$, $i = 1, \dots, k$, and $j = 1, \dots, k + l$, $\rho_{k, k} = 1$,

$$\rho_{k, k+l} := \lim_{n \rightarrow \infty} \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb} \left(\Lambda_p^{k+l} \cap \Lambda_m^k \right)^2, \quad (33)$$

and

$$\Psi((k, i), (k+l, j)) := \Sigma_{k, k+l}^0 \Sigma_{i, j}^0 + \Sigma_{k, j}^0 \Sigma_{i, k+l}^0, \quad (34)$$

with

$$\Lambda_m^k = (t_{m-1}^k, t_m^k], \quad k = 1, \dots, d \text{ and } m = 1, \dots, N_k, \quad (35)$$

and $\text{Leb}((a, b]) = b - a$ for $a < b$ and zero otherwise.

Remark 3 Note that the results in the previous theorem are in agreement with those derived in Hayashi and Yoshida (2008) and Christensen et al. (2013). We would like to emphasize that the innovation in our result relies on the joint convergence of the CholCov estimator under different asymptotic rates.

Remark 4 Within the framework of Assumption 1, the previous theorem can be generalized to the context of random sample times. Indeed, if we assume that the times on \mathcal{T}_k are random but independent of our driver Brownian motion, then the estimation vector β (as defined in Theorem 1) is asymptotically mixed Gaussian, providing that the following limit exists and is finite

$$\rho_{k,k+l} = \mathbb{P}\text{-}\lim_{n \rightarrow \infty} \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k)^2.$$

Remark 5 All our asymptotic results presented so far assume that σ is constant. When σ is time varying, relation (10) does not hold anymore, and the proposed CholCov estimator is no longer the MLE. If we assume that σ is not constant yet deterministic, and the CholCov is still constructed as in (16) and (17), then it follows from the proof of Proposition 1, that the k -th row of this misspecified CholCov estimator corresponds to the k -th row of the QMLE estimator obtained by maximizing (24) w.r.t. $\Sigma^0(k)$. Moreover, due to Xiu (2010), c.f. Aït-Sahalia et al. (2010), such a QMLE estimator is consistent and asymptotically normal, and under Assumption 1, it is asymptotically equivalent to the realized quadratic covariation computed on the mesh \mathcal{T}_k . This implies that when σ is time-varying, Proposition 2 remains true.

On the other hand, by reasoning as in the proof of Theorem 1, when σ is time-varying, yet deterministic, it holds that

$$\mathbb{E}[\beta_k(i) \beta_{k+l}(j)] = \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \int_{\Lambda_p^{k+l} \cap \Lambda_m^k} \int_{\Lambda_p^{k+l} \cap \Lambda_m^k} \Psi_{s,r}((k, i), (k+l, j)) ds dr, \quad (36)$$

where β_k as in Theorem 1 and

$$\Psi_{s,r}((k, i), (k + l, j)) := \Sigma_{k,k+l}^0(s) \Sigma_{i,j}^0(r) + \Sigma_{k,j}^0(s) \Sigma_{i,k+l}^0(r),$$

with $\Sigma^0(t) = \sigma_t \sigma_t'$. Hence, according to Theorem 2, a necessary and sufficient condition for the asymptotic normality of β is that the sequence in (36) converges to a finite number as $n \rightarrow \infty$. Under standard assumptions on σ , e.g. σ being bounded away from zero, we can verify, in the context of Assumption 1, that such a condition is satisfied.

Remark 6 Under the refresh-time sampling approach used throughout the paper, it is, for each grid k , the least liquid asset that determines the number of observations in that grid (N_k). A clear example where a violation of Assumption 1 invalidates the central limit distribution in Theorem 1 is the case where $\lim_{n \rightarrow \infty} N_{k+l}/N_k = 0$, for some integer $0 < l \leq (d - k)$. Indeed, setting $N_k = n^{\alpha_k}$ and $N_{k+l} = n^{\alpha_{k+l}}$, with $\alpha_k > 3\alpha_{k+l}$, then $\sqrt{N_k N_{k+l}}/N_{k+l}^2 \rightarrow \infty$. On the other hand, as shown in the proof of Theorem 1,

$$\mathbb{E}[\beta_k(i) \beta_{k+l}(j)] = \Psi((k, i), (k + l, j)) \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2.$$

Moreover, from equation (56) in the Appendix, if $\Lambda_p^{k+l} \cap \Lambda_m^k \neq \emptyset$

$$\sqrt{N_k N_{k+l}} Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2 = \frac{\sqrt{N_k N_{k+l}}}{4} [\Delta_m^k + \Delta_p^{k+l} - (|t_m^k - t_p^{k+l}| + |t_{m-1}^k - t_{p-1}^{k+l}|)]^2.$$

Thus, as $n \rightarrow \infty$

$$\sqrt{N_k N_{k+l}} Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2 > (\sqrt{N_k N_{k+l}}/N_{k+l}^2) (N_{k+l}^2 \Delta_p^{k+l})^2 \rightarrow \infty,$$

or in other words, the covariance between $\beta_k(i)$ and $\beta_{k+l}(j)$ does not converge. Hence, in this case Theorem 1 fails.

Remark 7 A feasible estimator of the asymptotic covariance matrix. Theorem 1 shows that the distribution of CholCov using (16) and (17) as candidate estimator has the asymptotic distribution of RCOV when data is synchronous. In case of non-synchronous prices, the asymptotic covariance of CholCov estimates corresponds to the covariance of RCOV multiplied by the scaling factor $\rho_{k,k+l}$.

This factor can be estimated using $\rho_{k,k+l}(n)$ computed on the observation times actually used in the CholCov estimation. For the estimation of Ψ , the asymptotic covariance of the realized covariance estimator, we recommend to use the estimator of Barndorff-Nielsen and Shephard (2004a). We suggest simply synchronizing all series at once to estimate this quantity, and will show in the simulations that this feasible estimator works well in finite samples.

Finally, note that Theorem 1 provides the asymptotic distribution of CholCov, which is a function of \hat{q} , i.e., the estimate of q^0 . Applying the delta method, we obtain the following asymptotic normality result for \hat{q} .

Corollary 1 *Let $q = (q_k)_{k=1,\dots,d}$, with $q_k = \sqrt{N_k} [\hat{q}_k - q_k^0]$ as in Proposition 2. Under Assumption 1, q is asymptotically normal with covariance structure given by*

$$\lim_{n \rightarrow \infty} \mathbb{E}[qq'] = \nabla q^{0'} \varphi \nabla q^0,$$

where ∇q^0 is the gradient of q^0 with respect to $\text{vech}(\Sigma^0)$ and φ is the asymptotic covariance matrix of the corresponding elements in $\beta = (\beta_k)_{k=1,\dots,d}$ as in Theorem 1.

5. Monte Carlo simulation

CholCov is positive semidefinite by construction and, in the presence of noise, we recommend using noise-robust sequential estimators of the underlying q^0 parameter such as the pre-averaging MRC estimator. In this section, we first use a Monte Carlo analysis to compare the accuracy of *mrcCholCov* using pre-averaging based estimates of q^0 with the standard MRC estimator of Christensen et al. (2010) and its composite counterpart, denoted cMRC. The standard MRC is estimated on a fully synchronized sample with N observations. It is implemented with a window length of $\lfloor N^{1/2+0.1} \rfloor$ at which it is guaranteed to be PSD, but for which the convergence is only at the rate $N^{1/5}$. For the cMRC estimator in which elements are estimated on pairwise grids, we use the $N^{1/4}$ rate MRC estimator with window length $\lfloor N^{1/2} \rfloor$. We use the same efficient estimator for *mrcCholCov*[•], which will still result in a PSD estimate. We report the average condition number of the different estimators for a range of sample sizes and show CholCov systematically has a lower condition number than its counterparts. We conclude this section by studying the finite-sample accuracy of the normality approximation for the finite-sample distribution of the realized covariance based CholCov estimates (assuming prices are observed without microstructure noise), as described in Section 4.

5.1. Setup

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

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

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

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

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 used to restart the process each day at $\varrho(0) \sim \mathcal{N}(0, (-2(\beta)^2/\alpha)^{-1})$. The parameter choice ensures that $E\left(\int_0^1 \sigma^{(i)2}(u) du\right) = 1$. 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 $X^{(i)} = Y^{(i)} + \epsilon^{(i)}$ with

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

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

For each estimator, 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

$$\text{Bias}_{k,l} = \hat{\Sigma}_{k,l} - \Sigma_{k,l} \quad \text{and} \quad \text{RMSE}_{k,l} = \sqrt{(\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2}. \quad (37)$$

We report the averages of the above two statistics over 1,000 replications. We consider average bias and RMSE over multiple elements, making a distinction between liquid and illiquid assets.

5.2. Comparison of CholCov with the standard (composite) MRC estimators

The comparison is done for $d = 20$ series with $\lambda = \{5, 5, \dots, 5, 120\}$.⁵ 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. (2015), 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 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 slightly over 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 1 reports the results of the simulation. Panel A reports the fraction of PSD estimates for the various estimates. The CholCov and MRC estimators are both PSD by construction. The cMRC is not ensured to be PSD, and the simulations show that in the vast majority of cases it is not.

Panel B reports estimation results of the various individual elements. We show the average bias and RMSE of all covariance, correlation and variance elements, as well as their averages for those elements involving the illiquid asset and those that do not separately. 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 larger for all elements.

Both versions of CholCov and cMRC are much more accurate. The only problem occurs in the variance estimate of the illiquid series, which is common to both estimators. The Strip-and-Replace estimation reduces RMSE, but has slightly larger bias compared to mrcCholCov. The bias in variances

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

transfers to a bias in the covariances involving the illiquid assets, but the correlation is estimated accurately. On average, $\text{mrcCholCov}^\bullet$ estimates the correlations with the least bias and the lowest RMSE, amongst the estimators considered.

Finally, we report more results on the quality of the estimates by reporting condition numbers for various dimensions. The condition number is defined as the ratio of the highest over the smallest eigenvalue. A problem is said to be ill-conditioned if this ratio is large. In practice, this is mainly due to extremely low eigenvalues. A well-conditioned matrix ensures that operations like inverting the matrix do not amplify estimation error, which is critical in many applications. As before, we investigate this within our simulations, where prices are simulated using poisson processes with parameters $\lambda = \{5, 5, \dots, 5, 120\}$, and we set the noise variance to $\xi^2 = 0.001$.

Table 2 reports the average condition number over 1,000 simulations of the true matrix ICOV, and the estimated matrices. It becomes clear that CholCov leads to slightly better conditioned matrices than cMRC and vastly better conditioned matrices than MRC. This is especially true for larger dimensional problems.

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, CholCov is PSD by construction, whereas the Composite estimator needs a regularization to become PSD.

5.3. Normality of the CholCov estimator in finite samples with stochastic volatility.

The final simulations aim validating the asymptotic theory derived in Section 4 for the realized covariance based CholCov estimator using the expanding refresh-time grids \mathcal{T}_k . We assess the robustness of the results to deviations from the assumptions on the underlying process. To do so, we simulate a five-dimensional process according to the DGP described in Subsection 5.1. We consider different assumptions on the DGP. First, we distinguish between constant and time-varying volatility. Second, we consider different types of asynchronicity. Type I has no asynchronicity, and all observations are equispaced. Type II has asynchronicity, where the first series is observed every second, 5% of data is randomly discarded for the second series and an additional 5% of data is randomly discarded for the each additional series. The series are asynchronous, but $T_{k+1} \subset T_k$. We present the rejection frequencies of the hypothesis of multivariate normality CholCov estimates using the squared

Mahalanobis Distance (MD)

$$(\text{rvech } CholCov - \text{rvech } ICOV)' \tilde{\varphi}^{-1} (\text{rvech } CholCov - \text{rvech } ICOV) \sim \chi_{d(d+1)/2}^2, \quad (38)$$

where $\tilde{\varphi}$ is the asymptotic covariance matrix of $\text{rvech } CholCov$, constructed using (32). We use infeasible and feasible estimates of φ , and consider 80% and 95% significance tests. For the infeasible version of φ we simply compute its population quantity using the simulated path, while for the feasible version we use the estimator given in Remark 7.

The results reported in Table 3 confirm the theory of the previous section, and show that relaxing the constant volatility assumption does not greatly affect the $CholCov$'s distribution. The rejection frequencies confirm that when there is no asynchronicity, i.e. Type I, the $RCOV$ asymptotic distribution is valid. Second, if there is asynchronicity of Type II, the elements converge at different rates based on the effective number of observations.

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.

We contribute to the multivariate approach to portfolio VaR analysis by studying 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 $mrcCholCov^\bullet$ estimates. We also consider $RCOV$ 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.⁶ The analysis uses one-second 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). 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. 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 retained 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, which is no longer invertible. Second, our application, like many applications involving covariance estimates, requires the estimate to be positive semidefinite, invalidating the use of the composite estimation technique.

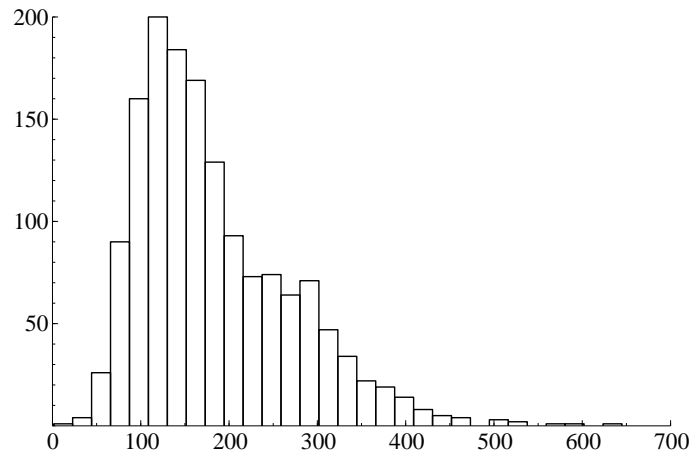
6.2. Methodology

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:

$$VaR_t^c = w_t' \mu_{t|t-1} + z_c \sqrt{w_t' S_{t|t-1} w_t}, \quad (39)$$

⁶The 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.



where $\mu_{t|t-1}$ is the vector conditional means and $S_{t|t-1}$ the conditional covariance matrix. Under the assumption of conditional normality, z_c is the c 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 $c = \{0.01, 0.025, 0.05, 0.95, 0.975, 0.99\}$. For long positions, the risk comes from a drop in the price and therefore an observation in the left tail of the distribution, whereas for short positions the right tail is important. The Value-at-Risk is for long positions when $c < 0.5$ and for short positions when $c > 0.5$.

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

CholCov and RCOV are ex-post measures, while for estimating the *VaR* a covariance forecast and therefore a dynamic model is needed. 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 Noureldin 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 to 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 α and β take the form of

$$S_{t|t-1} = (1 - \alpha - \beta)\Omega + \alpha V_{t-1} + \beta S_{t-1|t-2}, \quad (40)$$

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 covariance estimate based on intraday data. To reduce the number of parameters to be estimated, we apply covariance targeting, where Ω 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 in Equation (11) of Noureldin et al. (2012).

The cDCC models take the following form:

$$\begin{aligned} S_{t|t-1} &= D_{t|t-1} R_{t|t-1} D_{t|t-1} \\ R_{t|t-1} &= \text{diag}(Q_{t|t-1})^{-1/2} Q_{t|t-1} \text{diag}(Q_{t|t-1})^{-1/2} \\ Q_{t|t-1} &= (1 - \alpha - \beta)\bar{Q} + \alpha P_{t|t-1}^* + \beta Q_{t-1|t-2}, \end{aligned} \quad (41)$$

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 . For the univariate models we use ARFIMA(1, d , 0) models on the natural logarithm of the estimated variances, as there is ample evidence (e.g. Andersen et al., 2003) for the presence of long memory in realized variances. By modeling the natural logarithm of the variances, we ensure positive out-of-sample forecasts.⁷ 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 large-dimensional problems. We assume Gaussian innovations for the computation of the *VaR* for all

⁷We adjust for the bias caused by the log-transformation as in Giot and Laurent (2004).

models.

To obtain the forecasts we estimate all the models on an increasing window of observations, making one-step-ahead forecasts for the 1,000 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 :

$$Hit_t(c) = \begin{cases} 1 - c & \text{if } w'r_t < VaR_{t|t-1}^c \\ -c & \text{otherwise.} \end{cases} \quad (42)$$

Similarly, $Hit_t(1-c) = 1-c$ if $w'r_t > VaR_{t|t-1}^{1-c}$. We run the regressions $Hit_t(c) = \delta_0 + \sum_{k=1}^K \delta_k Hit_{t-k}(c) + \varepsilon_t$ and test the joint hypothesis $H_0 : \delta_0 = \delta_1 = \dots = \delta_K = 0, \forall k = 1, \dots, K$. *VaR* violations are uncorrelated over time if the δ_i with $i > 0$ are 0, whereas the unconditional coverage is correct if $\delta_0 = 0$. Denote by $\theta = (\delta_0, \delta_1, \dots, \delta_K)'$ the vector of parameters of the model and by X the matrix of explanatory variables of the regression. The test statistic is

$$\frac{\hat{\theta}' X' X \hat{\theta}}{c(1-c)}, \quad (43)$$

and follows a χ^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:

$$LF_t^c = (VaR_{t|t-1}^c - w'r_t) Hit_t(c). \quad (44)$$

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. We set the level to 90%, such that models are removed from the set until the final model set includes the best models with 90% confidence.⁸

⁸The MCS is carried out using the MulCom 3.0 package for Ox (Hansen and Lunde, 2014), using 10,000 resamples and a bootstrap block-length of 2.

6.3. Results

Table 4 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.

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 of the dynamic quantile test for all c and K 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 $c = 1\%$ or $c = 2.5\%$. 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 applied to RCOV offers some slight improvements over the daily Scalar-BEKK. However, the model is still almost uniformly rejected. Applying the HEAVY model to CholCov offers more substantial improvements. However, the short memory HEAVY model is still rejected by the data. For CholCov, conditional gaussianity is rejected for lag-length $K = 2$ but not for $K = 1$. This suggests more lags are needed, and as such we have considered a HEAVY(2,2) model, which is also rejected by the data.

The cRDCC with ARFIMA dynamics on the variances takes into account the long-memory properties of Realized Variance. The cRDCC model utilizing RCOV has similar performance to the cDCC model using daily returns, with rejections in the left tail. However, due to the flexibility of the model, combined with the accurate CholCov estimates based on liquidity sorted assets, it can 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 for $K = 2$.

To see the economic significance of using liquidity sorting instead of observation count sorting, Table 4 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 5 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 Cholcov estimator to make efficient use of the information content of high-frequency data and an appropriate dynamic model.

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 of 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

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

where the covariance forecast $S_{t|t-1}$ is partitioned into the firm k and the ten ETFs E :

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

We use a regression-based comparison that was proposed by Engle (2014). To do so, 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:

$$r_{k,t} = \alpha + \delta_k^{CholCov} Z_{k,t|t-1}^{CholCov} + \delta_k^{RCOV} Z_{k,t|t-1}^{RCOV} + \delta_k^{DCC} Z_{k,t|t-1}^{DCC} + \eta_{k,t}, \quad (45)$$

for each $k = 1, \dots, 52$. A perfect specification for the β -vector would lead to a δ coefficient equal to one, while the other coefficients would be zero. Engle (2014) does an in-sample comparison of a static beta with a DCC beta, and Hansen et al. (2014) add the Realized Beta GARCH to the analysis and perform an out-of-sample analysis.

It follows that 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 CholCov over RCOV. To test this, we consider the regression (45) 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 6. 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 δ 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 for over 40 percent of the assets. It almost completely encompasses the explanatory power of beta forecasts based on 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 as the superior model.

In the direct comparison between CholCov and RCOV forecasts, 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. Cholcov is flexible and can use any other estimator for the intermediate calculations, adopting their robustness properties. The asymptotic distribution is derived for the constant volatility, no-noise case, but with asynchronicity. Simulations confirm its properties and demonstrate that the resulting estimates are accurate and better conditioned than full-dimensional or composite estimates.

In an empirical application we use 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 is in contrast to models based on daily returns and dynamic models on 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. Acknowledgments

This research was financially supported by a grant of the Netherlands Organization for Scientific Research (NWO). This work was granted access to the HPC resources of Aix-Marseille Université financed by the project Equip@Meso (ANR-10-EQPX-29-01) of the program “Investissements d’Avenir” supervised by the Agence Nationale de la Recherche. 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, Maastricht University Seminar, Netherlands Econometric Study Group 2013, the 13th OxMetrics User Conference, the 7th International MIFN workshop, the 14th OxMetrics User Conference, the Conférence Séries Temporelles, Économétrie et Finance, and the 7th Annual SoFiE conference. The authors thank Francesco Violante for sharing his code for the RDCC model. Finally, we thank Yacine Aït-Sahalia (the editor), the associate editor, two anonymous referees, Eric Beutner, Jia Li, Mark Podolskij and Kevin Sheppard for useful comments.

10. References

- Aielli, G. P., 2013. Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* 31 (3), 282–299.
- 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., 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.
- Bibinger, M., Hautsch, N., Malec, P., Reiß, M., 2014. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *The Annals of Statistics* 42 (4), 1312–1346.
- Bibinger, M., Reiß, M., 2014. Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics* 41 (1), 23–50.
- 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 (1), 116–133.
- Christensen, K., Podolskij, M., Vetter, M., 2013. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *Journal of Multivariate Analysis* 120, 59–84.
- Corsi, F., Peluso, S., Audrino, F., 2015. Missing in asynchronicity: A Kalman-EM approach for multivariate realized covariance estimation. *Journal of Applied Econometrics* 30 (3), 377–397.

- De la Vega, W. F., 1974. On almost sure convergence of quadratic brownian variation. *The Annals of Probability* 2 (3), 551–552.
- Doornik, J., 2009. *Object-oriented Matrix Programming Using Ox*. Timberlake Consultants Press.
- Engle, R., 2014. Dynamic conditional beta. Working Paper.
- Engle, R., Shephard, N., Sheppard, K., 2008. Fitting vast dimensional time-varying covariance models. Working Paper.
- 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., 2014. Mulcom 3.00, an Ox software package for multiple comparisons.
- 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., 2014. Realized beta GARCH: Multivariate garch model with realized measures of volatility and covolatility. *Journal of Applied Econometrics* 29 (5), 774–799.
- 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 (4), 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 (4), 627–645.
- Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.
- Hayashi, T., Yoshida, N., 2008. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics* 60 (2), 367–406.
- Koike, Y., 2013. Estimation of integrated covariances in the simultaneous presence of nonsynchronicity, microstructure noise and jumps. *Econometric Theory*, 1–79.
- Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10 (5), 603–621.
- Lunde, A., Shephard, N., Sheppard, K., 2015. Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, forthcoming.
- Mykland, P. A., 2010. A gaussian calculus for inference from high frequency data. *Annals of Finance* 8 (2), 235–258.
- Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 27 (6), 907–933.
- Nualart, D., 2006. *The Malliavin Calculus and Related Topics*, 2nd Edition. Probability and its Applications. Springer Berlin Heidelberg.
- Nualart, D., Peccati, G., 2005. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* 33 (1), 177–193.
- Palandri, A., 2009. Sequential conditional correlations: Inference and evaluation. *Journal of Econometrics* 153, 122–132.
- Peccati, G., Tudor, C. A., 2005. Gaussian limits for vector-valued multiple stochastic integrals. In: *Séminaire de*

- Probabilités XXXVIII. Springer, pp. 247–262.
- Peluso, S., Corsi, F., Mira, A., 2014. A Bayesian high-frequency estimator of the multivariate covariance of noisy and asynchronous returns. *Journal of Financial Econometrics* 13 (3), 665–697.
- 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.
- 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* 48, 817–838.
- Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159 (1), 235–250.
- Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160 (1), 33–47.

Table 1: Results Simulation II

Panel A	mrcCholCov		mrcCholCov [*]		MRC		cMRC		PSD cMRC	
<i>Fraction PSD</i>										
$\xi^2 = 0.000$	1.000		1.000		1.000		0.377		1.000	
$\xi^2 = 0.001$	1.000		1.000		1.000		0.199		1.000	
$\xi^2 = 0.010$	1.000		1.000		1.000		0.000		1.000	
Panel B	mrcCholCov		mrcCholCov [*]		MRC		cMRC		PSD cMRC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
<i>Covariances</i>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.005	0.093	-0.003	0.091	0.025	0.230	-0.004	0.097	-0.004	0.097
$\xi^2 = 0.001$	-0.006	0.094	-0.005	0.093	0.030	0.233	-0.004	0.098	-0.004	0.098
$\xi^2 = 0.010$	-0.005	0.100	-0.007	0.096	0.035	0.246	-0.006	0.105	-0.006	0.105
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.002	0.086	-0.001	0.084	0.026	0.230	-0.001	0.090	-0.001	0.090
$\xi^2 = 0.001$	-0.004	0.088	-0.001	0.087	0.029	0.233	-0.001	0.091	-0.001	0.091
$\xi^2 = 0.010$	-0.002	0.093	-0.003	0.090	0.035	0.247	-0.003	0.098	-0.003	0.098
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.035	0.143	-0.034	0.142	0.025	0.226	-0.035	0.143	-0.035	0.143
$\xi^2 = 0.001$	-0.028	0.139	-0.036	0.135	0.029	0.230	-0.034	0.145	-0.034	0.144
$\xi^2 = 0.010$	-0.036	0.146	-0.041	0.137	0.034	0.241	-0.034	0.157	-0.034	0.154
<i>Correlations</i>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.002	0.019	-0.002	0.019	-0.004	0.037	0.000	0.020	-0.002	0.019
$\xi^2 = 0.001$	-0.002	0.019	-0.001	0.018	-0.004	0.039	-0.001	0.021	-0.002	0.020
$\xi^2 = 0.010$	0.001	0.026	-0.001	0.023	-0.008	0.060	-0.003	0.030	-0.010	0.030
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.001	0.017	-0.000	0.017	-0.004	0.037	0.000	0.017	-0.001	0.017
$\xi^2 = 0.001$	-0.002	0.017	-0.001	0.016	-0.004	0.039	-0.001	0.018	-0.001	0.018
$\xi^2 = 0.010$	0.001	0.025	-0.001	0.022	-0.008	0.060	-0.002	0.026	-0.007	0.026
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.012	0.032	-0.017	0.032	-0.004	0.035	-0.004	0.035	-0.006	0.034
$\xi^2 = 0.001$	-0.003	0.029	-0.002	0.028	-0.004	0.036	-0.004	0.036	-0.008	0.035
$\xi^2 = 0.010$	-0.001	0.035	0.001	0.031	-0.007	0.056	-0.007	0.056	-0.032	0.053
<i>Variances</i>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.004	0.111	-0.004	0.104	0.029	0.248	-0.004	0.104	-0.004	0.104
$\xi^2 = 0.001$	-0.004	0.114	-0.004	0.106	0.033	0.252	-0.004	0.106	-0.003	0.106
$\xi^2 = 0.010$	-0.004	0.129	-0.004	0.117	0.046	0.274	-0.004	0.117	-0.004	0.116
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.002	0.103	-0.001	0.097	0.029	0.249	-0.001	0.097	-0.001	0.097
$\xi^2 = 0.001$	-0.002	0.106	-0.001	0.098	0.033	0.253	-0.001	0.098	-0.001	0.098
$\xi^2 = 0.010$	-0.003	0.121	-0.002	0.109	0.046	0.275	-0.002	0.109	0.004	0.109
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.048	0.209	-0.058	0.199	0.026	0.227	-0.058	0.199	-0.055	0.198
$\xi^2 = 0.001$	-0.035	0.211	-0.055	0.204	0.033	0.232	-0.055	0.204	-0.049	0.202
$\xi^2 = 0.010$	-0.031	0.231	-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 the fraction of PSD estimates. 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.

Table 2: Results Simulation II: Condition numbers

d	ICOV	mrcCholCov	mrcCholCov [•]	MRC	cMRC
2	21.22	23.78	24.03	22.08	22.53
5	51.55	93.01	91.58	134.65	99.30
10	102.11	321.16	305.51	822.71	400.43
20	203.22	1892.58	1714.24	13878.14	2259.41
50	506.55	9978.12	9663.10	46336.40	10831.52

Note: Simulation results of the multivariate factor diffusion with $d = 2, 5, 10, 20, 50$ with $\lambda = 5, 5, \dots, 5, 120$ and $\xi^2 = 0.001$.

Table 3: Rejection rates of the normality assumption for the realized covariances based CholCov estimates

Asynchronicity Type	Constant Volatility	Noise	Infeasible		Feasible	
			80%	95%	80%	95%
I	Yes	No	19.8	4.9	19.1	4.3
II	Yes	No	18.7	5.2	17.0	3.9
I	No	No	20.8	4.1	18.5	5.8
II	No	No	19.5	4.2	18.9	5.7

Note: Type I has no asynchronicity, and all observations are equispaced. Type II has asynchronicity, where for the first series each second is observed, and for each next series, an additional 5% of observed prices are removed. The series are asynchronous, but $T_{k+1} \subset T_k$.

Table 4: Dynamic Quantile Test P-Values

	K\c	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 <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 <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 <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 <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 $\text{mrcCholCov}^{\bullet}$ covariance estimate.

Table 5: Loss function MCS results

c	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 (44).

Table 6: Beta Comparisons

<i>Panel A:</i> Summary of parameter estimates								
	Full Model			Excluding DCC		Single series		
	$\delta_i^{CholCov}$	δ_i^{RCOV}	δ_i^{DCC}	$\delta_i^{CholCov}$	δ_i^{RCOV}	$\delta_i^{CholCov}$	δ_i^{RCOV}	δ_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
<i>Panel B:</i> Rejection Frequencies								
$H_{CholCov} : 0.577$ $H_{RCOV} : 1.000$ $H_{DCC} : 0.962$ $H'_{CholCov} : 0.404$ $H'_{RCOV} : 1.000$								
<i>Panel C:</i> 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 (45). 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.

Appendix A: The Peccati and Nualart's fourth moment theorem

In this appendix, we present a version of Peccati and Nualart's fourth moment theorem (see Nualart and Peccati, 2005 and Peccati and Tudor, 2005) which plays a key role in the proof of Theorem 1.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $(W_t)_{0 \leq t \leq 1}$ a d -dimensional Brownian motion satisfying that $\mathcal{F} = \sigma\left((W_t)_{0 \leq t \leq 1}\right)$, the σ -algebra generated by $(W_t)_{0 \leq t \leq 1}$. Recall that the underlying log-price process Y_t is assumed to follow a multivariate Brownian Semi-Martingale with constant covariance, i.e.,

$$dY_t = \sigma_0 dW_t.$$

Observe that, since $\sigma^0 \sigma^{0r}$ is assumed to be positive definite, then $\mathcal{F} = \sigma\left((Y_t)_{0 \leq t \leq 1}\right)$.

Let $T = [0, 1] \times \{1, \dots, d\}$, and $\mu(dsdx) = ds\#(dx)$, where $\#(dx)$ is the counting measure on $\{1, \dots, d\}$. Define

$$I_1^W(\xi) := \sum_{i=1}^d \int_0^1 \xi(s, i) dW_s^i \quad (46)$$

$$I_2^W(\varphi) := 2 \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \int_0^r \varphi[(s, i), (r, j)] dW_s^i dW_r^j, \quad (47)$$

with $\xi : T \rightarrow \mathbb{R}$ μ -square integrable, and $\varphi : T \times T \rightarrow \mathbb{R}$ $(\mu \times \mu)$ -square integrable such that

$$\varphi[(s, i), (r, j)] = \varphi[(r, j), (s, i)], \quad (r, j), (s, i) \in T. \quad (48)$$

We define in an analogous way I_1^Y and I_2^Y . The collection of $(\mu \times \mu)$ -square integrable functions satisfying (48) will be denoted by $\mathcal{H}^{\odot 2}$.

The following result is an ad hoc simplification of Nualart and Peccati (2005) and Peccati and Tudor (2005):

Theorem 2 (Peccati and Tudor, 2005) *Let $m \geq 2$ and consider a sequence of functions $\{\varphi_n^1, \dots, \varphi_n^m\}_{n \geq 1}$ such that $\varphi_n^i \in \mathcal{H}^{\odot 2}$ and for every $i, j = 1, \dots, m$, the limit*

$$\lim_{n \rightarrow \infty} \mathbb{E} [I_2^W(\varphi_n^i) I_2^W(\varphi_n^j)] = C(i, j),$$

exists. Then the following statements are equivalent

1. The vector $(I_2^W(\varphi_n^i))_{i=1}^m$ converges in distribution to a m -dimensional Gaussian vector $N =$

$(N_1, \dots, N_m)'$ with covariance matrix $C = (C(i, j))_{i, j=1, \dots, m}$;

2. For every $j = 1, \dots, m$, $I_2^W(\varphi_n^j) \xrightarrow{d} N_j$.

Additionally, for our proof of Theorem 1, we will need the following result, obtained under the assumptions listed in this appendix.

Proposition 3

1. We have

$$I_1^Y(\xi^1) I_1^Y(\xi^2) = I_2^W(\sigma^0 \xi^1 \tilde{\otimes} \sigma^0 \xi^2) + \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \xi^1(s, i) \Sigma_{i,j}^0 \xi^2(s, j) ds, \quad (49)$$

where $\Sigma^0 = \sigma^0 \sigma^{0'}$, $\sigma^0 \xi(s, i) = \sum_{x=1}^d \sigma_{i,x}^0 \xi(s, x)$ and

$$\xi^1 \tilde{\otimes} \xi^2 [(s, i), (r, j)] = \frac{1}{2} [\xi^1(s, i) \xi^2(r, j) + \xi^1(r, j) \xi^2(s, i)],$$

i.e., the symmetric tensor product.

2. For every $\varphi^1, \varphi^2 \in \mathcal{H}^{\odot 2}$

$$\mathbb{E} [I_2^W(\varphi^1) I_2^W(\varphi^2)] = 2 \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \int_0^1 \varphi^1[(r, j), (s, i)] \varphi^2[(r, j), (s, i)] ds dr. \quad (50)$$

Proof. A general proof for part 2 can be found in Nualart (2006), so let us focus on the first part.

Since

$$dY_s^i = \sum_{j=1}^d \sigma_{ij}^0 dW_s^j,$$

we get, once we apply Itô's formula, that

$$\begin{aligned}
I_1^Y(\xi^1) I_1^Y(\xi^2) &= \sum_{i,j,x,y=1}^d \int_0^1 \xi^1(s,i) \sigma_{ij}^0 dW_s^j \int_0^1 \xi^2(s,x) \sigma_{xy}^0 dW_s^y \\
&= \sum_{i,j,x,y=1}^d \int_0^1 \int_0^s \xi^1(r,i) \sigma_{ij}^0 \xi^2(s,x) \sigma_{xy}^0 dW_r^j dW_s^y \\
&\quad + \sum_{i,j,x,y=1}^d \int_0^1 \int_0^s \xi^1(s,i) \sigma_{ij}^0 \xi^2(r,x) \sigma_{xy}^0 dW_r^y dW_s^j \\
&\quad + \sum_{i,j,x,y=1}^d \int_0^1 \xi^1(s,i) \sigma_{ij}^0 \xi^2(s,x) \sigma_{xy}^0 d[W^j, W^y]_s \\
&= \sum_{i,j=1}^d \int_0^1 \int_0^s \sigma^0 \xi^1(r,j) \sigma^0 \xi^2(s,i) dW_r^j dW_s^i \\
&\quad + \sum_{i,j=1}^d \int_0^1 \int_0^s \sigma^0 \xi^1(s,i) \sigma^0 \xi^2(r,j) dW_r^j dW_s^i \\
&\quad + \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \xi^1(s,i) \Sigma_{i,j}^0 \xi^2(s,j) ds.
\end{aligned}$$

Equation (49) follows by comparing the previous equation and (47). ■

Appendix B: Proofs

Proof of Proposition 1. In view that the mapping $(H, G) \mapsto HGH^T$ is a continuously differentiable bijection between $\mathbb{M}_{d, chol}^+$ and \mathbb{M}_d^+ and in this case $\widehat{\Sigma}$ is the realized covariance, we only need to check that $(\widehat{H}, \widehat{G}) = (\overline{H}, \overline{G})$. We only consider the bivariate case but the general case follows by induction. Trivially, we have that $\hat{g}_{11} = \bar{g}_{11}$. Now, since

$$\begin{aligned}
\hat{h}_{21} &= \frac{\sum_{j=1}^n r_l^{(2)}(T) f_j^{(1)}(T)}{\hat{g}_{11}} \\
&= \frac{\sum_{j=1}^n r_l^{(2)}(T) r_j^{(1)}(T)}{\hat{g}_{11}},
\end{aligned}$$

we obtain that $\hat{h}_{21} = \bar{h}_{21}$, or in other words $\widehat{H} = \overline{H}$. Finally, in view that for $i = 1, \dots, n$

$$\bar{f}_j^{(2)}(T) = r_j^{(2)}(T) - \frac{\bar{\Sigma}_{21}}{\bar{\Sigma}_{11}} r_j^{(1)}(T),$$

we get

$$\begin{aligned}
\hat{g}_{22} &= \sum_{j=1}^n \left[\bar{f}_j^{(2)}(T) \right]^2 \\
&= \sum_{j=1}^n \left[r_j^{(2)}(T) - \frac{\bar{\Sigma}_{21}}{\bar{\Sigma}_{11}} r_j^{(1)}(T) \right]^2 \\
&= \bar{\Sigma}_{22} - 2 \frac{\bar{\Sigma}_{21}}{\bar{\Sigma}_{11}} \sum_{j=1}^n r_j^{(1)}(T) r_j^{(2)}(T) + \left(\frac{\bar{\Sigma}_{21}}{\bar{\Sigma}_{11}} \right)^2 \sum_{i=1}^n \left[r_j^{(1)}(T) \right]^2 \\
&= \bar{\Sigma}_{22} - \frac{(\bar{\Sigma}_{21})^2}{\bar{\Sigma}_{11}} = \bar{g}_{22},
\end{aligned}$$

which concludes the proof. ■

Proof of Proposition 2. Following the same reasoning as in Remark 1, we have that

$$\left\| \widehat{\Sigma}(k) - \widetilde{\Sigma}(k) \right\| \leq \sup \left| \frac{1}{\Delta_j(T_k) N_k} - 1 \right| \left\| \widehat{\Sigma}(k) \right\|, \quad (51)$$

where $\widehat{\Sigma}(k)$ is the realized covariance of the process $\pi_k(X_t)$. Note that from (22) and (23), we get

$$\widehat{\Sigma}(k) - \Sigma^0(k) = \Pi_k \left(\widehat{\Sigma} - \Sigma^0 \right) \Pi_k',$$

and

$$\text{rvech} \left(\widehat{\Sigma}(k) - \Sigma^0(k) \right) = B_k \text{rvech} \left(\widehat{\Sigma} - \Sigma^0 \right),$$

for some matrix B_k . It follows that $\left\| \widehat{\Sigma}(k) - \widetilde{\Sigma}(k) \right\| = o_{\mathbb{P}} \left(N_k(n)^{-1/2} \right)$. The first conclusion of this proposition follows trivially from this and the fact that

$$q(k) - q^0(k) = \Pi_{k(k+1)/2} \left(\widehat{q} - q^0 \right),$$

with \widehat{q} as in Proposition 1. The final result follows from the following equation

$$\hat{q}_k - q_k^0 = A_k \left[\hat{q}(k) - q^0(k) \right], \quad k = 1, \dots, d.$$

■

Proof of Theorem 1. From Proposition 2, β_k converges to a normal random vector, for $k = 1, \dots, d$. Thus, for any $k = 1, \dots, d$ and $i = 1, \dots, k$, $\beta_k(i)$ is asymptotically normal. To show the desired

result of joint asymptotic normality, we use Theorem 2. Therefore, in order to conclude the proof, we only need to show that $\beta_k(i) = I_2^W(\varphi)$ for some $\varphi \in \mathcal{H}^{\odot 2}$ depending on (i, k) , and that the limit $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i) \beta_{k+l}(j)]$ exists and is given by (32), for every $k = 1, \dots, d$, $l = 1, \dots, d - k$, $i = 1, \dots, k$, and $j = 1, \dots, k + l$.

Let us first verify that $\beta_k(i) = I_2^W(\varphi)$. Let $\widehat{\Sigma}_{ij}^k$ be the ij th element of $\widehat{\Sigma}(k)$. Then

$$\begin{aligned} \widehat{\Sigma}_{ij}^k &= \sum_{l=1}^{N_k} \left(Y_{t_l^k}^i - Y_{t_{l-1}^k}^i \right) \left(Y_{t_l^k}^j - Y_{t_{l-1}^k}^j \right) \\ &= \sum_{l=1}^{N_k} I_1^Y[\xi^{i,l}(T_k)] I_1^Y[\xi^{j,l}(T_k)], \end{aligned}$$

where

$$\xi^{i,l}(T_k)(s, x) := \mathbf{1}_{\Lambda_l^k}(s) \delta_i(x), \quad i = 1, \dots, k,$$

with Λ_l^k as in (35) and δ_i the Dirac delta on i . Using (49) in Proposition 3, we have that

$$\begin{aligned} I_1^Y[\xi^{i,l}(T_k)] I_1^Y[\xi^{j,l}(T_k)] &= I_2^W(\sigma^0 \xi^{i,l}(T_k) \widetilde{\otimes} \sigma^0 \xi^{j,l}(T_k)) \\ &\quad + \sum_{x=1}^d \sum_{y=1}^d \int_0^1 \xi^{i,l}(T_k)(s, x) \Sigma_{i,j}^0 \xi^{j,l}(T_k)(s, y) ds, \end{aligned}$$

where Σ_{ij}^0 is the ij th element of Σ^0 . In view that

$$\sum_{x=1}^d \sum_{y=1}^d \int_0^1 \xi^{i,l}(T_k)(s, x) \Sigma_{x,y}^0 \xi^{j,l}(T_k)(s, y) ds = \Delta_l(T_k) \Sigma_{ij}^0,$$

we deduce that

$$\begin{aligned} \beta_k(i) &= \sqrt{N_k} \left(\widehat{\Sigma}_{ki}^k - \Sigma_{ki}^0 \right) \\ &= \sqrt{N_k} \sum_{l=1}^{N_k} I_2^W(\sigma^0 \xi^{k,l}(T_k) \widetilde{\otimes} \sigma^0 \xi^{i,l}(T_k)). \end{aligned} \tag{52}$$

We conclude that $\beta_k(i) = I_2^W(\varphi)$, with $\varphi = \sqrt{N_k} \sum_{l=1}^{N_k} (\sigma^0 \xi^{k,l}(T_k) \widetilde{\otimes} \sigma^0 \xi^{i,l}(T_k)) \in \mathcal{H}^{\odot 2}$.

We now show that the limit $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i) \beta_{k+l}(j)]$ exists and is given by (32). From (52), for

every $k = 1, \dots, d$, $i = 1, \dots, k$, $l = 1, \dots, d - k$, $j = 1, \dots, k + l$.

$$\begin{aligned} \mathbb{E} [\beta_k(i) \beta_{k+l}(j)] &= \sqrt{N_k N_{k+l}} \mathbb{E} \left\{ \left[\sum_{m=1}^{N_k} I_2^W \left(\sigma^0 \xi^{k,m}(T_k) \tilde{\otimes} \sigma^0 \xi^{i,m}(T_k) \right) \right] \left[\sum_{p=1}^{N_{k+l}} I_2^W \left(\sigma^0 \xi^{k+l,p}(T_{k+l}) \tilde{\otimes} \sigma^0 \xi^{j,p}(T_{k+l}) \right) \right] \right\} \\ &= \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \mathbb{E} \left[I_2^W \left(\sigma^0 \xi^{k,m}(T_k) \tilde{\otimes} \sigma^0 \xi^{i,m}(T_k) \right) I_2^W \left(\sigma^0 \xi^{k+l,p}(T_{k+l}) \tilde{\otimes} \sigma^0 \xi^{j,p}(T_{k+l}) \right) \right]. \quad (53) \end{aligned}$$

Furthermore, since

$$\sigma^0 \xi^{k,m}(T_k) \tilde{\otimes} \sigma^0 \xi^{i,m}(T_k) [(s, x), (r, y)] = \frac{1}{2} \mathbf{1}_{\Lambda_m^k}(s) \mathbf{1}_{\Lambda_m^k}(r) (\sigma_{kx}^0 \sigma_{iy}^0 + \sigma_{ky}^0 \sigma_{ix}^0),$$

we get from (50)

$$\mathbb{E} \left[I_2^W \left(\sigma^0 \xi^{k,m}(T_k) \tilde{\otimes} \sigma^0 \xi^{i,m}(T_k) \right) I_2^W \left(\sigma^0 \xi^{k+l,p}(T_{k+l}) \tilde{\otimes} \sigma^0 \xi^{j,p}(T_{k+l}) \right) \right] = \Psi((k, i), (k+l, j)) \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k)^2,$$

with $\Psi((k, i), (k+l, j))$ as in (34). This implies that

$$\mathbb{E} [\beta_k(i) \beta_{k+l}(j)] = \Psi((k, i), (k+l, j)) \rho_{k,k+l}(n),$$

where

$$\rho_{k,k+l}(n) := \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k)^2.$$

If $\Psi((k, i), (k+l, j)) = 0$, the limit $\lim_{n \rightarrow \infty} E[\beta_k(i) \beta_{k+l}(j)]$ exists trivially, so without loss of generality we may and do assume that $\Psi((k, i), (k+l, j)) \neq 0$. Consequently, $\lim_{n \rightarrow \infty} E[\beta_k(i) \beta_{k+l}(j)]$ exists if and only if the limit $\lim_{n \rightarrow \infty} \rho_{k,k+l}(n)$ exists and it is finite. To prove this, we first prove in Lemma 2 that this limit in case of equispaced observations exists and is finite. We then show in Lemma 3 that, under Assumption 1, the error in approximating $\rho_{k,k+l}$ by its analog for equispaced observations is convergent. Together, these results imply that

$$\rho_{k,k+l}(n) = \tilde{\rho}_{k,k+l}(n) + O(1), \quad (54)$$

where

$$\tilde{\rho}_{k,k+l}(n) := \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb}(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k)^2,$$

with

$$\tilde{\Lambda}_m^k = \left(\frac{(m-1)}{N_k}, \frac{m}{N_k} \right], \quad m = 1, \dots, N_k.$$

Moreover, the error is convergent. Lemma 2 below guarantees that the limit of the sequence $\tilde{\rho}_{k,k+l}(n)$ exists and is finite, so $\rho_{k,k+l}(n)$, which concludes the proof. \blacksquare

Lemma 1 *Let*

$$\tilde{F}_n^{k,k+l} := \left\{ (m, p) : \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) > 0 \right\}.$$

where $\tilde{\Lambda}_m^k = \left(\frac{m-1}{N_k}, \frac{m}{N_k} \right]$. Then

$$\#\tilde{F}_n^{k,k+l} = N_k + O(N_k), \tag{55}$$

where $\#$ denotes the counting measure.

Proof. If $N_k = N_{k+l}$, the result is trivial, so without loss of generality, we may and do assume that $N_k > N_{k+l}$. Moreover, from Assumption 1, either $\lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} = 1$ or $0 < \lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} < 1$.

On the other hand, note that in general

$$\text{Leb} \left(\Lambda_p^{k+l} \cap \Lambda_m^k \right) = \max \left\{ t_m^k \wedge t_p^{k+l} - t_{m-1}^k \vee t_{p-1}^{k+l}, 0 \right\},$$

where \wedge and \vee stand respectively for the maximum and minimum of two numbers. Due to the formulas

$$a \wedge b = \frac{1}{2} (a + b - |a - b|); \quad c \vee d = \frac{1}{2} (c + d + |c - d|),$$

we deduce that

$$\text{Leb} \left(\Lambda_p^{k+l} \cap \Lambda_m^k \right) = \frac{1}{2} \max \left\{ \Delta_m^k + \Delta_p^{k+l} - (|t_m^k - t_p^{k+l}| + |t_{m-1}^k - t_{p-1}^{k+l}|), 0 \right\}, \tag{56}$$

and thus that

$$\text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) = \frac{1}{2} \max \left\{ \frac{1}{N_k} + \frac{1}{N_{k+l}} - \left(\left| \frac{m}{N_k} - \frac{p}{N_{k+l}} \right| + \left| \frac{m-1}{N_k} - \frac{p-1}{N_{k+l}} \right| \right), 0 \right\}.$$

This implies that $(m, p) \in \tilde{F}_n^{k,k+l}$ if and only if

$$\frac{1}{N_k} + \frac{1}{N_{k+l}} > \left| \frac{m}{N_k} - \frac{p}{N_{k+l}} \right| + \left| \frac{m-1}{N_k} - \frac{p-1}{N_{k+l}} \right|. \tag{57}$$

Now, decompose

$$\tilde{F}_n^{k,k+l} = \tilde{F}_n^{k,k+l} (1) \cup \tilde{F}_n^{k,k+l} (2) \cup \tilde{F}_n^{k,k+l} (3),$$

where

$$\begin{aligned} \tilde{F}_n^{k,k+l} (1) &:= \left\{ (m, p) : m = p, \text{Leb} \left(\tilde{\Lambda}_m^{k+l} \cap \tilde{\Lambda}_m^k \right) > 0 \right\}, \\ \tilde{F}_n^{k,k+l} (2) &:= \left\{ (m, p) : m < p, \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) > 0 \right\}, \\ \tilde{F}_n^{k,k+l} (3) &:= \left\{ (m, p) : m > p, \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) > 0 \right\} \end{aligned}$$

are mutually disjoint. As a consequence,

1. $(m, p) \in \tilde{F}_n^{k,k+l} (1)$ if and only if $m = p$ and

$$m(N_k - N_{k+l}) < N_k; \tag{58}$$

2. $(m, p) \in \tilde{F}_n^{k,k+l} (2)$ if and only if $m < p$ and

$$(p - m) N_k + m(N_k - N_{k+l}) < N_k; \tag{59}$$

3. $(m, p) \in \tilde{F}_n^{k,k+l} (3)$ if and only if $m > p$ and one of the following holds:

- (a) $m < \frac{N_k}{N_{k+l}} p$ and

$$(p - m) N_{k+l} + p(N_k - N_{k+l}) < N_k; \tag{60}$$

- (b) $m \geq \frac{N_k}{N_{k+l}} p$ and

$$(m - p) N_{k+l} + p(N_{k+l} - N_k) < N_{k+l}. \tag{61}$$

As we will see later, the cardinality of $\tilde{F}_n^{k,k+l} (1)$, $\tilde{F}_n^{k,k+l} (2)$ and $\tilde{F}_n^{k,k+l} (3)$ depends on the limiting behavior of $\frac{N_k}{N_{k+l}}$, reason why we divide our analysis of the cardinality of $\tilde{F}_n^{k,k+l}$ in two cases.

Let $(m, p) \in \tilde{F}_n^{k,k+l}$. Then:

Case 1. If $\lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} = 1$, we have that one and only one of the following holds:

1. $(m, p) \in \tilde{F}_n^{k,k+l} (1)$ or equivalently from (58) and the fact that $N_k = N_{k+l} + o(N_k)$ with $o(N_k) > 0$, $m = p$ and

$$m < \frac{N_k}{o(N_k)},$$

or in other words

$$\frac{N_k}{o(N_k)} - 1 \leq \#\tilde{F}_n^{k,k+l}(1) < \frac{N_k}{o(N_k)}.$$

In view that

$$\limsup_{n \rightarrow \infty} \left(1 - \frac{1}{o(N_k)}\right) = \begin{cases} 1 & \text{if } \liminf_{n \rightarrow \infty} o(N_k) = \infty \\ 1 - \frac{1}{\liminf_{n \rightarrow \infty} o(N_k)} & \text{if } \liminf_{n \rightarrow \infty} o(N_k) < \infty, \end{cases}$$

where we used the fact that $\infty \geq \liminf_{n \rightarrow \infty} o(N_k) = \liminf_{n \rightarrow \infty} (N_k - N_{k+l}) > 0$. We infer that $\limsup_{n \rightarrow \infty} \frac{1}{N_k} \left(N_k - \frac{N_k}{o(N_k)}\right) < \infty$, or in other words

$$\#\tilde{F}_n^{k,k+l}(1) = N_k + O(N_k).$$

2. $(m, p) \in \tilde{F}_n^{k,k+l}(2)$ which, from (59), holds if and only if $m < p$ and

$$\begin{aligned} 1 &> (p - m) + m \frac{1}{N_k} (N_k - N_{k+l}) \\ &= (p - m) + mo(1), \end{aligned}$$

i.e., $(p - 1)(1 + o(1)) < m < p$. This implies that, whenever $\lim_{n \rightarrow \infty} \left(\frac{N_k}{N_{k+l}} - 1\right) = 0$, $\tilde{F}_n^{k,k+l}(2)$ is empty for n large enough.

3. $(m, p) \in \tilde{F}_n^{k,k+l}(3)$, which by (60), (61), and the fact that $1 + o(1) = \frac{N_k}{N_{k+l}}$, corresponds to one and only one of the following situations:

3.1. Equation (60) holds and $p < m < p + o(1)$. Thus, as in the previous case, $\tilde{F}_n^{k,k+l}(3)$ is empty for n large enough.

3.2. It holds that

$$p(1 + o(1)) \leq m < p(1 + o(1)) + 1.$$

Once again, we conclude that for n large enough, $\tilde{F}_n^{k,k+l}(3)$ is empty.

As a consequence, we conclude that if $\lim_{n \rightarrow \infty} \left(\frac{N_k}{N_{k+l}} - 1\right) = 0$, then

$$\begin{aligned} \#\tilde{F}_n^{k,k+l} &= \#\tilde{F}_n^{k,k+l}(1) \\ &= N_k + O(N_k). \end{aligned}$$

Case 2. Now, suppose that $\frac{N_{k+l}}{N_k} = K_{k+l,k} + o(1)$, with $0 < K_{k+l,k} < 1$. Then, it holds that $\lim_{n \rightarrow \infty} \left(\frac{N_k}{N_k - N_{k+l}} \right) = (1 - K_{k+l,k})^{-1} =: K_{k,k+l}^*$. As in the previous case, $(m, p) \in \tilde{F}_n^{k,k+l}$ if and only if

1. $(m, p) \in \tilde{F}_n^{k,k+l}(1)$, i.e. $m = p$ and

$$m < \frac{N_k}{N_k - N_{k+l}}.$$

Consequently, $(m, m) \in \tilde{F}_n^{k,k+l}(1)$ if and only if $m < K_{k,k+l}^*$ which reads, for n large enough, as

$$\#\tilde{F}_n^{k,k+l}(1) = K_{k,k+l}^* + o(1).$$

2. $(m, p) \in \tilde{F}_n^{k,k+l}(2)$ which, from (59), holds if and only if $m < p$ and

$$\begin{aligned} 1 &> p - m \frac{N_{k+l}}{N_k} \\ &= p - m(K_{k+l,k} + o(1)); \end{aligned}$$

i.e. $(p-1)(K_{k+l,k} + o(1))^{-1} < m < p$. This implies that, for n large enough, $\#\tilde{F}_n^{k,k+l}(2)$ is empty.

3. By reasoning as in the previous case, if $(m, p) \in \tilde{F}_n^{k,k+l}(3)$, then:

$$3.1. (p-1) \left(K_{k+l,k}^{-1} + o(1) \right) < m < p \left(K_{k+l,k}^{-1} + o(1) \right);$$

$$3.2. p \left(K_{k+l,k}^{-1} + o(1) \right) \leq m < 1 + p \left(K_{k+l,k}^{-1} + o(1) \right).$$

We conclude that

$$\begin{aligned} \#\tilde{F}_n^{k,k+l}(3) &= \#\left\{ (m, p) : (p-1) \left(K_{k+l,k}^{-1} + o(1) \right) < m < p \left(K_{k+l,k}^{-1} + o(1) \right) \right\} \\ &= K_{k+l,k}^{-1} N_{k+l} + o(1) \\ &= N_k + O(N_k). \end{aligned}$$

Hence

$$\begin{aligned} \#\tilde{F}_n^{k,k+l} &= \#\tilde{F}_n^{k,k+l}(1) + \#\tilde{F}_n^{k,k+l}(3) \\ &= K_{k,k+l}^* + o(1) + N_k + O(N_k) \\ &= N_k + O(N_k). \end{aligned}$$

We see that in all cases, (55) holds. ■

Lemma 2 *Under the notation of the previous lemma, we have that*

$$\lim_{n \rightarrow \infty} \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right)^2 < \infty.$$

Proof. As in the proof of the previous lemma, we divide the analysis in two cases, namely $\lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} = 1$ or $0 < \lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} < 1$.

Case 1. If $\lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} = 1$, we have from the proof of the previous lemma that

$$\sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right)^2 = \sum_{(m,m) \in \tilde{F}_n^{k,k+l}(1)} Leb \left(\tilde{\Lambda}_m^{k+l} \cap \tilde{\Lambda}_m^k \right)^2.$$

Now, from (56), if $(m, m) \in \tilde{F}_n^{k,k+l}(1)$ then

$$N_k Leb \left(\tilde{\Lambda}_m^{k+l} \cap \tilde{\Lambda}_m^k \right) = \frac{N_k}{N_{k+l}} - m \left(\frac{N_k}{N_{k+l}} - 1 \right) \rightarrow 1, \quad \text{as } n \rightarrow \infty.$$

Therefore

$$\begin{aligned} \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right)^2 &= \sum_{(m,m) \in \tilde{F}_n^{k,k+l}(1)} \left(\frac{1}{N_k} + o(N_k^{-1}) \right)^2 \\ &= \sqrt{N_{k+l} N_k} (N_k + O(N_k)) \left(\frac{1}{N_k} + o(N_k^{-1}) \right)^2 \\ &= 1 + o(1). \end{aligned}$$

Case 2. By assuming that $\frac{N_{k+l}}{N_k} = K_{k+l,k} + o(1)$, with $0 < K_{k+l,k} < 1$ and using the proof of the previous lemma we have that if $(m, m) \in \tilde{F}_n^{k,k+l}(1)$, then $m < K_{k,k+l}^*$ and

$$N_k Leb \left(\tilde{\Lambda}_m^{k+l} \cap \tilde{\Lambda}_m^k \right) = \frac{N_k}{N_{k+l}} - m \left(\frac{N_k}{N_{k+l}} - 1 \right) \rightarrow K_{k+l,k}^{-1} - m \left(K_{k+l,k}^{-1} - 1 \right), \quad \text{as } n \rightarrow \infty.$$

Furthermore, if $(m, p) \in \tilde{F}_n^{k,k+l}(3)$ then $p < m < \frac{N_k}{N_{k+l}} p$, and

$$N_k Leb \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) = \frac{N_k}{N_{k+l}} - \left(\frac{N_k}{N_{k+l}} p - m \right) \rightarrow K_{k+l,k}^{-1} - \left(K_{k+l,k}^{-1} p - m \right), \quad \text{as } n \rightarrow \infty.$$

Using the previous asymptotics and applying Lemma 1, we conclude that

$$\begin{aligned} \sqrt{N_{k+l}N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right)^2 &= \sum_{(m,m) \in \tilde{F}_n^{k,k+l}} \left(\frac{1}{N_k} + O(N_k^{-1}) \right)^2 \\ &= \sqrt{N_{k+l}N_k} (N_k + O(N_k)) \left(\frac{1}{N_k} + O(N_k^{-1}) \right)^2 \\ &= \sqrt{\frac{N_{k+l}}{N_k}} + O(1), \end{aligned}$$

where the error term is convergent as $n \rightarrow \infty$.

To sum up, in all cases, the conclusion of this lemma holds. ■

Lemma 3 *Under Assumption 1, we have that*

1. For any $0 < \beta \leq 3$

$$\text{Leb} \left(\Lambda_p^{k+l} \cap \Lambda_m^k \right) = \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) + o \left[(N_k N_{k+l})^{-\beta} \right].$$

2. Denoting $\#$ the counting measure,

$$\#F_n^{k,l} = \#\tilde{F}_n^{k,l} + o(1),$$

with $\tilde{F}_n^{k,l}$ as in Lemma 1.

Proof. From (56) we get that

$$\begin{aligned} 2 \left| \text{Leb} \left(\Lambda_p^{k+l} \cap \Lambda_m^k \right) - \text{Leb} \left(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) \right| &\leq \left| \Delta_m^k - \frac{1}{N_k} \right| + \left| \Delta_p^{k+l} - \frac{1}{N_{k+l}} \right| \\ &+ \left| t_m^k - \frac{m}{N_k} \right| + \left| t_p^{k+l} - \frac{p}{N_{k+l}} \right| \\ &+ \left| t_{m-1}^k - \frac{m-1}{N_k} \right| + \left| t_{p-1}^{k+l} - \frac{p-1}{N_{k+l}} \right|. \end{aligned} \tag{62}$$

It is immediate from Assumption 1, that for any $0 < \beta \leq 3$, as $n \rightarrow \infty$

$$(N_k N_{k+l})^\beta \left| \Delta_m^k - \frac{1}{N_k} \right| = (N_{k+l}/N_k)^\beta N_k^{2\beta-1} |\Delta_m^k N_k - 1| \rightarrow 0.$$

On the other hand, once again, applying Assumption 1, we see that

$$\begin{aligned} (N_k N_{k+l})^\beta \left| t_m^k - \frac{m}{N_k} \right| &\leq (N_{k+l}/N_k)^\beta N_k^{2\beta-1} \sum_{j=1}^m |\Delta_j^k N_k - 1| \\ &\leq m(N_{k+l}/N_k)^\beta N_k^{2\beta-1} \sup_j |\Delta_j^k N_k - 1| \rightarrow 0, \quad n \rightarrow \infty. \end{aligned} \tag{63}$$

Part 1 is obtained as a combination of (62) and (63).

Now we proceed to verify part 2. To do this, it is enough to show that as $n \rightarrow \infty$,

$$\# \left(F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) + \# \left(\tilde{F}_n^{k,l} \setminus F_n^{k,l} \right) \rightarrow 0. \tag{64}$$

We only verify that $\# \left(F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) \rightarrow 0$, the other part can be obtained in an analogous way. To do this, we show that for n large, still finite, $F_n^{k,l} \setminus \tilde{F}_n^{k,l}$ is empty. Let us, as previously, divide our analysis in two cases, namely $\lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} = 1$ or $0 < \lim_{n \rightarrow \infty} \frac{N_{k+l}}{N_k} < 1$.

Case 1. If $N_k = N_{k+l} + o(N_k)$, we have from the proof of Lemma 1, that for n large enough,

$$F_n^{k,l} \setminus \tilde{F}_n^{k,l} = \left\{ (m, p) : m = p, \frac{N_k}{o(N_k)} < m \leq N_k, Leb(\Lambda_m^{k+l} \cap \Lambda_m^k) > 0 \right\}.$$

From the first part and (56), $(m, p) \in F_n^{k,l} \setminus \tilde{F}_n^{k,l}$ if and only if $\frac{N_k}{o(N_k)} < m \leq N_k$

$$0 < Leb(\Lambda_m^{k+l} \cap \Lambda_m^k) = \frac{1}{N_{k+l}} - m \left(\frac{1}{N_{k+l}} - \frac{1}{N_k} \right) + o \left[(N_k N_{k+l})^{-\beta} \right],$$

for some $\beta > 1$, which implies that

$$m < \frac{N_k}{o(N_k)} + o(1).$$

Therefore, $(m, p) \in F_n^{k,l} \setminus \tilde{F}_n^{k,l}$ if and only if

$$\frac{N_k}{o(N_k)} < m < \frac{N_k}{o(N_k)} + o(1).$$

Therefore, for n enough large, $F_n^{k,l} \setminus \tilde{F}_n^{k,l}$ is empty.

Case 2. Now, suppose that $\frac{N_{k+l}}{N_k} = K_{k+l,k} + o(1)$, with $0 < K_{k+l,k} < 1$. By the proof of Lemma 1,

for n large

$$F_n^{k,l} \setminus \tilde{F}_n^{k,l} = F_n^{k,l} \setminus \tilde{F}_n^{k,l} (1) \cup F_n^{k,l} \setminus \tilde{F}_n^{k,l} (3).$$

From the first part of this lemma, $(m, p) \in F_n^{k,l} \setminus \tilde{F}_n^{k,l} (1)$ if and only if

$$m < K_{k,k+l}^* + o(1),$$

where $K_{k+l,k}^* = (1 - K_{k+l,k})^{-1}$, meaning that we can choose n_0 finite, such that $F_n^{k,l} \setminus \tilde{F}_n^{k,l} (1)$ is empty for any $n \geq n_0$. Reasoning as in the proof of Lemma 1, we can, in an analogous way as before, deduce that $F_n^{k,l} \setminus \tilde{F}_n^{k,l} (3)$ is empty for n large enough.

To conclude, this implies that $\#(F_n^{k,l} \setminus \tilde{F}_n^{k,l}) = 0$ for n sufficiently large. ■

Appendix C: The score function and Fisher Information of the ML estimators under the assumptions of Proposition 1

This section is devoted to deriving the score function, the Hessian matrix, and the Fisher information for the ML estimator of q^0 under the assumptions of Proposition 1. For notational clarity we will omit the 0 subscript and the grid dependence of the durations in the remainder of this section. We derive these quantities with respect to generic g_{kk} and h_{kl} elements.

Remark 8 Note that $f^{(k)} = r^{(k)} - \sum_{l=1}^{k-1} h_{kl} f^{(l)}$ when $k > l$. As such, $\frac{\partial f^{(k)}}{\partial h_{lm}} = -\sum_{i=1}^{k-1} h_{ki} \frac{\partial f^{(i)}}{\partial h_{lm}}$ for $k > l > m$, and 0 otherwise.

We first compute the score functions:

$$\frac{\partial \bar{Q}_n}{\partial g_{kk}} = -\frac{1}{2} \sum_{j=1}^N \left[\frac{1}{\Delta_j g_{kk}} - \frac{f_j^{(k)2}}{\Delta_j g_{kk}^2} \right], \quad (65)$$

$$\frac{\partial \bar{Q}_n}{\partial h_{kl}} = -\frac{1}{2} \left[\sum_{i=1}^d \sum_{j=1}^N \frac{2f_j^{(i)}}{\Delta_j g_{ii}} \frac{\partial f_j^{(i)}}{\partial h_{kl}} \right]. \quad (66)$$

The Hessian is as follows:

$$\frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial g_{kk}} = \sum_{j=1}^N \left[\frac{1}{2\Delta_j g_{kk}^2} - \frac{f_j^{(k)2}}{\Delta_j g_{kk}^3} \right] \quad (67)$$

$$\frac{\partial^2 \bar{Q}_n}{\partial h_{mn} \partial g_{kk}} = \frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial h_{mn}} = - \sum_{j=1}^N \left[\frac{f_j^{(k)}}{\Delta_j g_{kk}^2} \frac{\partial f_j^{(k)}}{\partial h_{mn}} \right] \quad (68)$$

$$\frac{\partial^2 \bar{Q}_n}{\partial h_{kl} \partial h_{mn}} = - \sum_{j=1}^N \left[\sum_{i=1}^d \frac{1}{\Delta_j g_{ii}} \left[\frac{\partial f_j^{(i)}}{\partial h_{mn}} \frac{\partial f_j^{(i)}}{\partial h_{kl}} + f_j^{(i)} \frac{\partial^2 f_j^{(i)}}{\partial h_{kl} \partial h_{mn}} \right] \right]. \quad (69)$$

We compute the components of the Fisher matrix as the negative of the expected value of the Hessian.

$$-\mathbb{E} \left[\frac{\partial^2 \bar{Q}_n}{\partial g_{kk}^2} \right] = -\mathbb{E} \left[\sum_{j=1}^N \frac{1}{2\Delta_j g_{kk}^2} - \frac{f_{t_j}^{(k)2}}{g_{kk}^3} \right] = \frac{1}{2g_{kk}^2} \quad \forall k \quad (70)$$

$$-\mathbb{E} \left[\frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial h_{mn}} \right] = -\mathbb{E} \left[- \sum_{j=1}^N \left[\frac{f_{t_j}^{(k)}}{\Delta_j g_{kk}^2} \frac{\partial f_{t_j}^{(k)}}{\partial h_{mn}} \right] \right] = 0, \quad (71)$$

since not a single partial derivative of $f_j^{(k)}$ contains $f_j^{(k)}$ and $\mathbb{E}(f^{(k)} f^{(l)}) = 0$ for $k \neq l$.

Next let, $p = \max(k, m)$.

$$-\mathbb{E} \left[\frac{\partial^2 \bar{Q}_n}{\partial h_{kl} \partial h_{mn}} \right] = \begin{cases} 0 & \text{if } l \neq n \\ \frac{g_{pp}}{g_{ll}} \frac{\partial^2 f^{(p)}}{\partial h_{kl} \partial f^{(l)}} \frac{\partial^2 f^{(p)}}{\partial h_{mn} \partial f^{(l)}} & \text{otherwise,} \end{cases} \quad (72)$$

where the last two partial derivatives are in fact functionals of h_{kl} .