Abstract

We introduce LASSO–type regularization for large dimensional realized covariance estimators of log–prices. The procedure consists of shrinking the off–diagonal entries of the inverse realized covariance matrix towards zero. This technique produces covariance estimators that are positive definite and with a sparse inverse. We name the estimator realized network, since estimating a sparse inverse realized covariance matrix is equivalent to detecting the partial correlation network structure of the daily log-prices. The large sample consistency and selection properties of the estimator are established. An application to a panel of US bluechips shows the advantages of the estimator for out–of–sample GMV asset allocation.

Keywords: Networks, Realized Covariance, Lasso

JEL: C13, C33, C52, C58
1 Introduction

The covariance matrix of the log–prices of financial assets is a fundamental ingredient in many applications ranging from asset allocation to risk management. For more than a decade now the econometric literature has made a number of significant leaps forward in the estimation of covariance matrices using financial high frequency data. This new generation of estimators, commonly referred to as realized covariance estimators, measure precisely the daily covariance of log–prices using intra-daily price information. The literature has proposed an extensive number of procedures that allow to estimate the covariance efficiently under general assumptions, such as the presence of market microstructure noise and asynchronous trading in the data generating process.

Despite the significant leaps forward, the estimation of large realized covariance matrices has a number of hurdles. First, as it has been put forward by Hautsch, Kyj, and Oomen (2012) and Hautsch, Kyj, and Malec (2015), it is hard to estimate precisely the covariance matrix when the number of assets is large. Second, in large systems it is challenging to synthesize effectively the information contained in the covariance matrix and unveil the cross-sectional dependence structure of the data. In this work we propose a realized covariance estimation strategy that tackles simultaneously both of these challenges. The estimation approach consists of using LASSO–type shrinkage to regularize realized covariance estimators. The LASSO procedure detects and estimates the nonzero partial correlations among the daily log–prices. The set of nonzero partial correlations can then be represented as a network. Our proposed estimation approach has different highlights. If the partial correlation structure of the daily log–prices is sufficiently sparse, then the regularized estimator can deliver substantial accuracy gains over its unregularized counterpart. Moreover, detecting the network of interconnections among the daily log prices is interesting in the light of the recent strand of research on networks in economics by, among others, Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), which shows that in highly interconnected systems the most highly interconnected entities influence the aggregate behaviour of the entire system.

In its more general version, the framework we work in makes a number of fairly common assumptions on the dynamics of the asset prices (cf Bandi and Russell 2006; Ait-Sahalia, Mykland, and Zhang 2005; Fan, Li, and Yu 2012). We assume that observed log prices are equal to the efficient log prices, which are Brownian semi–martingales, plus a noise term that is due to
market microstructure frictions. Prices are observed according to the realization of a counting process driving the arrival of trades/quotes of each asset and are allowed to be asynchronous. The target estimation of interest is integrated covariance matrix of the efficient daily log–prices.

We introduce a network definition built upon the integrated covariance, which we call the integrated partial correlation network. Assets $i$ and $j$ are connected in the integrated partial correlation network iff the partial correlation between $i$ and $j$ implied by the integrated covariance is nonzero. As is well known, the network is entirely characterized by the inverse of the integrated covariance matrix, which we call hereafter the integrated concentration matrix. In fact, it has been known since at least Dempster (1972) that if the $(i,j)$–th entry of the inverse covariance matrix is zero, then variables $i$ and $j$ are partially uncorrelated, that is, are uncorrelated conditional on all other assets. Thus, the sparsity structure of the integrated concentration matrix determines the partial correlation network dependence structure among the daily log–prices.

We use lasso to obtain a sparse integrated concentration matrix estimator. The procedure consists of regularizing a consistent realized covariance estimator. Several realized covariance estimators have been introduced in the literature in the presence of market microstructure effects and asynchronous trading. In this work we focus in particular on the Two–Scales Realized Covariance estimators (TSRC) and the Multivariate Realized Kernel (MRK) based on pairwise refresh sampling (Ait-Sahalia et al. 2005; Barndorff-Nielsen, Hansen, Lunde, and Shephard 2011; Fan et al. 2012). These estimators are then regularized using the glasso (Friedman, Hastie, and Tibshirani 2011), which shrinks the off–diagonal elements of the inverse of the covariance estimators to zero. The procedure allows to detect the nonzero linkages of the integrated partial correlation network. Moreover, the sparse integrated concentration matrix estimator can be inverted to obtain an estimator of the integrated covariance.

We study the large sample properties of the realized network estimator, and establish conditions for consistent estimation of the integrated concentration and consistent selection of the integrated partial correlation network. We develop the theory for the TSRC and MRK estimators based on pairwise refresh–sampling built upon the general asymptotic theory developed by Ravikumar, Wainwright, Raskutti, and Yu (2011). The MRK estimator results are obtained by developing a novel concentration inequality, while for the TSRC estimator we apply a concentration inequality derived in Fan et al. (2012). Results are established in a high–dimensional setting, that is, we allow for the total number of parameters to be larger than the amount of observations available to the extent that the proportion of nonzero parameters is small relative
to the total. Other realized covariance estimators satisfying an appropriate concentration assumption lead to regularized estimators with similar properties. We emphasize that the theory is developed under essentially the same assumptions for both estimators. In particular, we assume the market microstructure noise that contaminates prices is iid. It is important to stress however that the MRK can also handle more complex types of market microstructure dependence but that we do not take advantage of this in this current work.

A simulation study is used to investigate the finite sample properties of the procedure. Different specifications of the integrated covariance matrix of the efficient price process are used to assess the precision of the realized network estimator. The procedure is also benchmarked against a set of alternative regularization techniques proposed in the literature, including shrinkage (Ledoit and Wolf, 2004) and factor based approaches. Among others results, simulations show that when the integrated concentration matrix is indeed sparse the realized network achieves the best performance among the set of candidate regularization procedures we consider.

We apply the realized network methodology to analyse the network structure of a panel of US bluechips throughout 2009 using the TSRC, MRK as well as the classic Realized Covariance (RC) estimators. More precisely, we use the realized network to regularize what we call idiosyncratic realized covariance matrix, that is the residual covariance matrix of the assets after netting out the influence of the market factor. Results show that after controlling for the market factor, assets still exhibit a significant amount of cross-sectional dependence. The estimated networks are indeed sparse, with the number of estimated links being roughly 5% of the total possible number of linkages. The distribution of the connections of the assets exhibits power law behavior, that is the number of connections is heterogeneous and the most interconnected stocks have a large number of connections relative to the total number of links. The stocks in the industrial and energy sectors show a high degree of sectoral clustering, that is there is a large number of connections among firms in these industry groups. Technology companies and Google in particular are the most highly interconnected firms throughout the year. We investigate the usefulness of our procedure from a forecasting perspective by carrying out a Markowitz type Global Minimum Variance (GMV) portfolio prediction exercise. We run a horse race among different (regularized) covariance estimators to assess which estimator produces GMV portfolio weights that deliver the minimum out-of-sample GMV portfolio variance. Results show that the realized network significantly improves prediction accuracy irrespective of the covariance estimator used.

The rest of the paper is structured as follows. In Section 2 we introduce the base framework and the realized network estimator. The theoretical properties of the estimation procedure are analysed in Section 3. Section 4 introduces a number of important extensions to the baseline framework. Section 5 contains a simulation exercise to study the properties of the realized network estimator. Section 6 presents an application to a panel of US bluechips. Concluding remarks follow in Section 7.

2 Methodology

In this section we introduce the baseline framework and estimation approach. Important extensions of the baseline methodology, including allowing for market microstructure frictions, are considered later in Section 4.
2.1 Model

Let \( y(t) \) denote the \( n \)-dimensional log–price vector of \( n \) assets at time \( t \). We assume that the dynamics of \( y(t) \) are given by

\[
y(t) = \int_0^t b(u) \, du + \int_0^t \Theta(u) \, dB(u) \quad , \quad t \in [0, 1],
\]

where \( B(t) \) is an \( n \)-dimensional Brownian motion on a filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})\).

The drift \( b(t) \) is an \( n \)-dimensional process, and the spot covolatility process \( \Theta(t) \) is an \( n \times n \) positive definite random matrix. The entries of \( b(t) \) and \( \Theta(t) \) are assumed to be predictable and uniformly bounded on \([0, 1]\), and those of \( \Theta(t) \) are also assumed to be càdlàg.

We are considering our process over a fixed time interval of length 1, which typically represents a day. We set \( y = y(1) \). Also, we define the \( i \)–th entry of \( y \) as \( y_i \), for \( i = 1, \ldots, n \). One of the main objects of interest in this work is the ex–post covariance matrix of \( y \), which is given by

\[
\text{Var} (y) = \int_0^1 \Sigma(t) \, dt ,
\]

where \( \Sigma(t) = \Theta(t)\Theta(t)' = (\sigma_{ij}(t)) \) is the spot covariance matrix. We define the integrated covariance as \( \Sigma^* = \int_0^1 \Sigma(t) \, dt = (\sigma^*_{ij}) \).

In this work we introduce a network representation of \( y \) based on the partial linear dependence structure induced by the integrated covariance matrix. We call this network the integrated partial correlation network. Recall that a network is defined as an undirected graph \( G = (V, E) \), where \( V \) is the set of vertices \( V = \{1, 2, \ldots, n\} \) and \( E \) is the set of edges \( E \subseteq V \times V \). In the integrated partial correlation network the set of vertices corresponds to the set of assets, and a pair of assets is connected by an edge iff their daily returns are partially correlated given all other daily returns in the panel, that is,

\[
E = \{ (i, j) \in V \times V , \rho^{ij} \neq 0 , i \neq j \} ,
\]

where \( \rho^{ij} \) is the integrated partial correlation, a measure of linear cross-sectional conditional dependence defined as

\[
\rho^{ij} = \text{Cor}(y_i, y_j | \{y_k, k \neq i, j\}) .
\]

The integrated partial correlations can be characterized by the inverse integrated covariance ma-
matrix $\mathbf{K}^* = (\Sigma^*)^{-1} = (k^*_{ij})$ \cite{dempster1972, lauritzen1996}, which we call hereafter integrated concentration matrix. We recall that

$$\rho_{ij} = \frac{-k^*_{ij}}{\sqrt{k^*_{ii}k^*_{jj}}}.$$  

Thus, the integrated partial correlation network can be equivalently defined as

$$\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V}, k^*_{ij} \neq 0, i \neq j\}.$$  

It is important to emphasize that the integrated partial correlation network definition represents particular correlation relations among the daily log–prices. Obviously enough, the absence of correlation between the log–daily prices of two assets does not necessarily imply that the spot prices are also uncorrelated.

### 2.2 Estimation

We are interested in (i) estimating the integrated covariance and concentration matrices of the daily log–prices, and (ii) detecting the nonzero entries of the integrated concentration matrix. The estimation strategy we follow consists of applying LASSO type regularization on the standard realized covariance estimators proposed in the literature.

We assume that the log-prices $y_i(t)$ of all assets $i = 1, \ldots, n$, are discretely observed at a same time grid $T = \{t_1, t_2, \ldots, t_m\}$ where $t_0 = 0 < t_1 < \cdots < t_m = 1$. We consider a generic estimator of the integrated covariance $\Sigma^*$ denoted $\overline{\Sigma} = (\overline{\sigma}_{ij})$ based on the observations $y_i(t_{\ell})$, $i = 1, \ldots, n$, $\ell = 1, \ldots, m$. We assume that this estimator satisfies the following concentration inequality.

**Assumption 1.** There exist positive constants $a_1, a_2$ and $a_3$ such that for all $i, j \in \{1, \ldots, n\}$, $x \in [0, a_1]$, and $m$ large,

$$P \left( |\overline{\sigma}_{ij} - \sigma^*_{ij}| > x \right) \leq a_2 m^{a_0} \exp(-a_3 (m^\beta x)^a),$$  

for some positive exponents $\beta, a$ and $a_0 \in \{0, 1\}$.

A natural estimator of the integrated covariance of $y$ in this setting is the so called Realized Covariance (RC) estimator. This estimator is the multivariate extension of the realized variance,
whose working mechanism is that the quadratic variation of the univariate price process can be
approximated by the sum of squared returns over small intervals.

**Realized Covariance Estimator.** The realized covariance estimator $\Sigma_{RC}$ is defined as

$$
\sigma_{RC,ij} = \sum_{k=1}^{m} (y_{ik} - y_{ik-1}) (y_{jk} - y_{jk-1}),
$$

where $y_{ik} = y_i(t_k)$.

Assume that the time grid satisfies that

$$
\sup_{\ell \in \{1, \ldots, m\}} |t_{\ell} - t_{\ell-1}| \leq \frac{c}{m},
$$

for some constant $c > 0$. Then Barndorff-Nielsen and Shephard (2004b) shows that the difference between $\sigma_{RC,ij}$ and $\sigma_{ij}^*$ is asymptotically normal with mean zero and variance of order $O \left( \frac{1}{m} \right)$.

Also, it is proved in Fan et al. (2012, Lemma 3), that under the conditions of this Section the estimator satisfies Assumption [1] with $\alpha_0 = 0, a = 2$ and $\beta = \frac{1}{2}$.

Given an estimator of the integrated covariance $\Sigma$ satisfying Assumption [1], we use the Graphical lasso procedure (glasso) to estimate the integrated concentration matrix $K^*$. 

**Realized Network Estimator.** Let $\hat{\Sigma}$ be an estimator of the integrated covariance, then we define the realized network estimator of the integrated concentration matrix as

$$
\hat{K}_{\lambda} = \arg \min_{K \in S^n} \left\{ \text{tr}(\Sigma K) - \log \det(K) + \lambda \sum_{i \neq j} |k_{ij}| \right\},
$$

where $\lambda \geq 0$ is the glasso tuning parameter and $S^n$ is the set of $n \times n$ symmetric positive definite matrices. The entries of $\hat{K}_{\lambda}$ are denoted by $(\hat{k}_{\lambda,ij})$. The corresponding realized covariance estimator based on the realized network is $\hat{\Sigma}_{\lambda} = \hat{K}_{\lambda}^{-1}$.

Observe that (4) defines a shrinkage type estimator. If we set $\lambda = 0$ in (4), we obtain the normal log-likelihood function of the covariance matrix, which is minimized by the inverse realized covariance estimator $(\Sigma)^{-1}$. If $\lambda$ is positive, (4) becomes a penalized likelihood function with penalty equal to the sum of the absolute values of the non-diagonal entries in the estimator.

The important feature of the absolute value penalty is that for $\lambda > 0$, some of the entries of the realized network estimator are going to be set to zero. The highlight of this estimator
is that it simultaneously estimates and selects the nonzero entries of $K^*$, thus providing an estimate of the linkages in the network. For this reason we dub the estimator realized network estimator. Banerjee and Ghaoui (2008) show that the optimization problem in (4) can be solved through a series of LASSO regressions, which motivates an iterative algorithm to solve (4) given in Friedman, Hastie, Hofling, and Tibshirani (2007). The highlight of the procedure is that it is straightforward to carry out the minimization of equation (4) even when the number of series is large. Importantly, the algorithm is also guaranteed to provide a positive definite estimate of the concentration matrix provided that the initial value of the algorithm is a positive definite matrix. Moreover, the algorithm only requires the $\Sigma$ estimator to be positive semi–definite (provided that $\lambda$ is larger than zero).

In order to apply the estimator in empirical applications we need to use a selection criterion to pick the value of the tuning parameter $\lambda$. In this work we resort to a BIC–type criterion defined as

$$BIC(\lambda) = m \times \left[ -\log \det \hat{K}_\lambda + \text{tr}(\hat{K}_\lambda \Sigma) \right] + \log m \times \#\{(i,j) : 1 \leq i < j \leq n, \hat{K}_{\lambda ij} \neq 0\},$$

as it is suggested in, among others, Yuan and Lin (2007).

3 Theory

In this section, we apply the theory of Ravikumar et al. (2011) to our particular case of an exponential concentration inequality to establish the large sample properties of the realized network estimator defined in (4).

In order to state the results we need to adopt the following notations. Given a matrix $U = (u_{ij}) \in \mathbb{R}^{\ell \times m}$, we set $||U||_\infty$, $||U||_1$, and $|||U|||_\infty$ to denote $\max_{i,j} |u_{ij}|$, $\sum_{i,j} |u_{ij}|$, and $\max \sum_{k=1}^m |u_{jk}|$, where $i \in \{1, 2, \ldots, \ell\}$ and $j \in \{1, 2, \ldots, m\}$. If $A = (a_{ij})$ is a $p \times q$ matrix and $B$ is an $m \times n$ matrix, the Kronecker product of matrices $A$ and $B$ is the $pm \times qn$ matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{bmatrix}.$$
We index the \( pm \) rows of \( A \otimes B \) by

\[
R = \{(1,1), (2,1), \ldots, (m,1), (1,2), (2,2), \ldots, (m,2), \ldots, (1,p), \ldots, (m,p)\},
\]

and the \( qn \) columns by

\[
C = \{(1,1), (2,1), \ldots, (n,1), (1,2), (2,2), \ldots, (n,2), \ldots, (1,q), \ldots, (n,q)\}.
\]

For any two subsets \( \bar{R} \subset R \) and \( \bar{C} \subset C \), we denote by \( (A \otimes B)_{\bar{R} \bar{C}} \) the matrix such that

\[
(A \otimes B)_{\bar{R} \bar{C}}(i,j)(c,d) \text{ is an entry of } (A \otimes B)_{\bar{R} \bar{C}} \iff (i,j) \in \bar{R} \text{ and } (c,d) \in \bar{C}.
\]

Assumption 2. Consider the \( n^2 \times n^2 \) matrix \( \Gamma^* = \Sigma^* \otimes \Sigma^* \). There exists some \( \alpha \in (0,1] \) such that

\[
\max_{e \in S^c} ||\Gamma^*_{eS} (\Gamma^*_S S)^{-1}||_1 \leq 1 - \alpha,
\]

where \( S = \mathcal{E} \cup \{(i,i) | i \in \mathcal{V}\} \) and \( S^c = \{(i,j) \in \mathcal{V} \times \mathcal{V}, k_{ij}^* = 0\} \).

Assumption 2 limits the amount of dependence between the non-edge terms (indexed by \( S^c \)) and the edge-based terms (indexed by \( S \)). The limit is controlled by \( \alpha \): the bigger the \( \alpha \) the smaller the dependence. In other words, if we set

\[
X_{(j,k)} = y_j y_k - \mathbb{E}(y_j y_k), \quad \text{for all } j, k \in \mathcal{V},
\]

then the correlation between \( X_{(j,k)} \) and \( X_{(\ell,m)} \) is low for any \( (j,k) \in S \) and \( (\ell,m) \in S^c \).

In the following, Theorem 1 shows that: (a) the rate at which the realized network estimator converges to the true value as the sample size \( m \) increases, and (b) a lower bound on the probability of correctly detecting the nonzero partial correlations (as well as their signs) as a function of the sample size \( m \). In particular, the estimator is model selection consistent with high probability, when \( n \) is large.

**Theorem 1.** Assume Assumptions 1 and 2 hold, and choose \( \lambda = \frac{8}{\alpha} m^{-\beta} \left( \log(a_2 n^{a_0 n^\tau}) \right)^{-\frac{1}{\beta}} \) in (4), where \( \tau > 2 \) is arbitrary.

(a) Assume that

\[
m > \left( \frac{2^{a_0}}{a_3} \log \left( a_2 n^\tau \left( a_3^{-\frac{1}{a_0}} c_0^{\frac{a_0}{a_1}} \right) \right) \right)^{-\frac{1}{a_0}} c_0^{\frac{1}{a_1}}, \tag{5}
\]
where
\[
\begin{align*}
c_0 &:= \max \left( \frac{1}{a_1}, 6(1 + 8\alpha^{-1})d \max(C_{\Sigma^*}, C_{\Sigma^*}^2, C_{\Gamma^*}), \frac{\alpha_0}{a_3} \exp \left( \frac{2}{a_2 \beta} \right) 1_{\{\alpha_0 = 1\}}, \frac{1}{\sigma_m} \right). 
\end{align*}
\]

Here, \(\sigma_m = \min_i \sigma_{ii}^*\), \(d\) is the maximum degree of the network, that is, the maximum number of edges that include a vertex, and we have set \(C_{\Gamma^*} = \|{(\Gamma^*)^{-1}}\|_{\infty}\) and \(C_{\Sigma^*} = \|\Sigma^*\|_{\infty}\).

Then,
\[
P \left( \|\hat{K}_\lambda - K^*\|_{\infty} \leq 2 \left( 1 + 8\alpha^{-1} \right) C_{\Gamma^*} m^{-\beta} \left( \frac{\log(a_2 m \alpha_0 n^\tau)}{a_3} \right)^{\frac{1}{\beta}} \right) \geq 1 - \frac{1}{n^{r-2}}. \tag{7}
\]

(b) Define \(\bar{c}_0 = \max \left( c_0, 2C_{\Gamma^*} \frac{(1 + 8\alpha^{-1})}{k_m} \right)\), where \(k_m\) is the minimum absolute value of the non-zero entries of \(K^*\). Assume that
\[
m > \left( \frac{2\alpha_0}{a_3} \log \left( a_2 n^\tau \left( \frac{a_3^{-\frac{1}{\beta}} \bar{c}_0^{\frac{1}{\beta}}}{\tau_0^{\frac{1}{\beta}}} \right)^{\alpha_0} \right) \right)^{\frac{1}{\beta}} \frac{1}{\tau_0}. 
\]

Then,
\[
P \left( \text{sign}(\hat{k}_{\lambda ij}) = \text{sign}(k_{ij}^*), \forall i, j \in V \right) \geq 1 - \frac{1}{n^{r-2}}. 
\]

Observe that when Assumption 1 holds with \(\alpha_0 = 0\), Theorem 1 is a direct application of Theorems 1 and 2 of Ravikumar et al. (2011). We give the proof of Theorem 1 for the case \(\alpha_0 = 1\) in the appendix.

The parameter \(d\) controls the degree of sparsity of our network. It ranges from 0 (fully sparse) to \(n\) (fully interconnected). To explore its effects, let us assume that the parameters \(C_{\Sigma^*}, C_{\Gamma^*}, \alpha\) and \(\sigma_m\) in Theorem 1 remain constant as a function \((n, m, d)\). In this case, when \(d = 0\), condition 5 means that \(m\) should not be smaller than \(O \left( (\log n)^{\frac{1}{\beta}} \right)\). When \(d = n\), condition 5 means that \(m\) should not be smaller than \(O \left( (\log n)^{\frac{1}{\beta}} n^{\frac{1}{\beta}} \right)\), since in this case \(c_0 = O(n)\). In other words, the sparser the network is, the lesser observations are required to estimate the concentration matrix accurately.
4 Extensions

4.1 Microstructure Noise and Asynchronicity

Rather than the efficient price, it is customary to assume that the econometrician observes the transaction (or midquote) price. This differs from the efficient price because trades (quotes) are affected by an array of market frictions that go under the umbrella term of market microstructure. Moreover, it is common to assume that the trades (quotes) of different assets are executed (posted) asynchronously. In this section we extend the baseline framework of Section 2 and introduce a number of realized covariance estimators designed to handle microstructure noise and asynchronous trading.

We assume that the log-prices of each asset $i$ are observed asynchronously on different time grids $T_i = \{t_{i1}, ..., t_{im_i}\}$, $i = 1, ..., n$. For each asset $i = 1, ..., n$ the econometrician observes the transaction (or midquote) prices $x_i(t_{i\ell})$ defined as

$$x_i(t_{i\ell}) = y_i(t_{i\ell}) + u_i(t_{i\ell}),$$

where $u_i(t_{i\ell})$ denotes the microstructure noise associated with the $\ell$th trade. We assume that the microstructure noise satisfies

$$u_i(t_{i\ell}) \sim i.i.d \ N(0, \eta_i^2),$$

where $\eta_i^2$ is some positive constant. We also assume that the noise process is independent of the efficient price process, and we use the same notation $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathcal{P})$ for the product filtered probability space of the two processes. Moreover the noise processes of different assets are independent of each other.

A standard technique used to handle asynchronous trading for realized covariance estimation is refresh time sampling, which was introduced by Barndorff-Nielsen et al. (2011). Several variants of this technique exist, like the pairwise and groupwise refresh time approaches, used in Fan et al. (2012), Lunde et al. (2011) and Hautsch et al. (2012). In this work we use pairwise refresh time sampling. Pairwise refresh time sampling based covariance estimation consists of estimating each entry of the covariance separately. The $i,j$–entry of the matrix is computed by first synchronizing the observations of assets $i$ and $j$ using refresh time and then estimating the covariance between assets $i$ and $j$ using the synchronized data. Notice that this approach does not guarantee that the covariance estimator is positive definite, as each covariance entry is
estimated using different subsets of observations.

More precisely, pairwise refresh time sampling is designed as follows. Consider two stocks $i$ and $j$ observed at times $T_i = \{t_{i1}, \ldots, t_{im}\}$ and $T_j = \{t_{j1}, \ldots, t_{jm}\}$ on $[0, 1]$. The set of pairwise refresh time points is $T_{ij} = \{\tau_0, \tau_1, \ldots, \tau_{m-1}, \tau_m, \tau_{m+1}\}$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_{m-1} < \tau_m \leq \tau_{m+1} = 1$ and $m$ is the amount of pairwise refresh time observations for stocks $i$ and $j$ on $[0, 1]$. For $0 < k \leq m$, $\tau_k$ is determined by

$$\tau_k = \max \{ \min \{ t \in T_i : t > \tau_{k-1} \}, \min \{ t \in T_j : t > \tau_{k-1} \} \}.$$ 

The timestamps for assets $i$ and $j$ that correspond to the refresh time $\tau_k$ are respectively $t_{i,k}^r = \max \{ t \in T_i : t \leq \tau_k \}$ and $t_{j,k}^r = \max \{ t \in T_j : t \leq \tau_k \}$. Accordingly, we obtain the refresh sample prices which are $x_{i,k}^r = \{x_i(t_{i1}^r), \ldots, x_i(t_{im}^r)\}$ for stock $i$ and $x_{j,k}^r = \{x_j(t_{j1}^r), \ldots, x_j(t_{jm}^r)\}$ for stock $j$. We use the shorthand notation $x_{l,k}^r$, $y_{l,k}^r$ and $u_{l,k}^r$ to denote $x_l(t_{l,k}^r)$, $y_l(t_{l,k}^r)$ and $u_l(t_{l,k}^r)$, respectively. Also, we define $M$ as the minimum pairwise refresh sample size across all pairs of assets.

After the data have been opportunely synchronized, a number of market microstructure noise robust estimators can be applied. First, we consider the Two–Scale Realized Covariance (TSRC) estimator proposed in Zhang, Mykland, and Ait-Sahalia (2005), which is a multivariate extension of the two–scale estimator.

**Two-scale Realized Covariance Estimator.** The Two-scale Realized Covariance Estimator $\Sigma_{TS}$ based on pairwise refresh time is defined as

$$\sigma_{TS,ij} = \frac{1}{K} \sum_{k=K+1}^{m} (x_{i,k}^r - x_{i,k-K}^r) (x_{j,k}^r - x_{j,k-K}^r) - \frac{m_K}{m_J} \sum_{k=J+1}^{m} (x_{i,k}^r - x_{i,k-J}^r) (x_{j,k}^r - x_{j,k-J}^r),$$

where $m_K = \frac{m-K+1}{K}$ and $m_J = \frac{m-J+1}{J}$.

Zhang et al. (2005) show that the optimal choice of $K$ has order $K = O(m^{\frac{2}{3}})$, and $J$ can be taken to be a constant such as 1. The first component of this estimator is the average of $K$ realized variances, and it converges to $\sigma_{ij}^r$ in the absence of noise. The second component is set to correct the bias caused by the noise. Under the optimal choice of $K$ and $J$, the estimation error is asymptotically normal with mean 0 and variance of order $O\left(m^{-\frac{2}{3}}\right)$. If we further assume that $\frac{1}{2}m^{\frac{1}{3}} \leq m_K \leq 2m^{\frac{1}{3}}$, then, under the condition that the synchronized observation
times satisfy condition,
\[
\sup_{\ell \in \{1, \ldots, m\}} |\tau_\ell - \tau_{\ell-1}| \leq \frac{c}{m},
\]
(8)

Fan et al. (2012) show that this estimator satisfies Assumption 1 with \(\beta = \frac{1}{6}, a = 2\) and \(\alpha_0 = 0\), and thus Theorem 1. Observe that since \(M\) is defined as the minimum pairwise refresh sample size across all pairs of assets, we should replace \(m\) with \(M\) when applying Theorem 1 to TSRC.

In the empirical implementation, for each pair of assets we choose \(J\) as 1 and \(K\) as the (rounded) average of the optimal bandwidth for the realized two scale volatility estimator of the two assets, following the procedure detailed in Ait-Sahalia et al. (2005).

Another robust estimator in the setting of this section is the multivariate realized kernel (MRK) estimator, which is the multivariate extension of the realized kernel estimator proposed in Barndorff-Nielsen et al. (2008). Kernel smoothing is a familiar technique in nonparametric econometrics, and MRK uses the properties of the kernel functions to shrink the impact of the noise.

**Multivariate Realized Kernel Estimator.** The multivariate realized kernel estimator \(\Sigma_K\) based on pairwise refresh time is defined as

\[
\sigma_{Kij} = \gamma_0 \left( x_{ri}, x_{rj} \right) + \frac{1}{2} \sum_{h=1}^{H} k \left( \frac{h-1}{H} \right) \left( \gamma_h(x_{ri}, x_{rj}) + \gamma_{-h}(x_{ri}, x_{rj}) \right),
\]

(9)

where for each \(h \in \{-H, -H+1, \ldots, -1, 0, 1, \ldots, H-1, H\}\),

\[
\gamma_h(x_{ri}, x_{rj}) = \sum_{p=1}^{m-1} (x_{ri p} - x_{ri p-1})(x_{rj p-h} - x_{rj p-h-1}).
\]

The kernel function \(k : [0,1] \to \mathbb{R}\) is assumed to be bounded and twice differentiable with bounded derivatives, and satisfies that \(k(1) = k'(0) = k'(1) = 0\) and \(k(0) = 1\).

For a choice of \(H\) of \(O(m^{\frac{1}{2}})\), Barndorff-Nielsen et al. (2008) show that the estimation error is asymptotically normal with mean 0 and variance of order \(O\left(m^{-\frac{1}{2}}\right)\). A typical choice of the kernel is the Parzen kernel, defined as,

\[
k(x) = \begin{cases} 
1 - 6x^2 + 6x^3 & 0 \leq x \leq \frac{1}{2} \\
2(1 - x)^3 & \frac{1}{2} < x \leq 1.
\end{cases}
\]

In the appendix, we show that this estimator follows the following concentration inequality.
Theorem 2. If the synchronized observation times satisfy condition (3), then there exist positive constants $a_1, a_2$ and $a_3$ such that for all $i, j \in \{1, ..., n\}$, $x \in [0, a_1]$, and $M$ large,

$$P \left( \left| \sigma_{K_{ij}} - \sigma_{ij}^* \right| > x \right) \leq a_2 M \exp(-a_3 M^{1/4} x).$$

Therefore, MRK satisfies Assumption 1 with $\beta = \frac{1}{4}$, $a = 1$ and $a_0 = 1$. Hence, we can apply Theorem 1 and we obtain that the estimation error of $\hat{K}_\lambda$ converges to zero at rate $m^{-\frac{1}{4}} \sqrt{\log n}$ (assuming that all other parameters including $\alpha, d, C_{\Gamma^*}$ and $C_{\Sigma^*}$ are constants). Thus, in this case, $m^\frac{1}{2}$ is required to be large compared to $\log n$ to make the error small in probability. Notice that this result is analogous to the one obtained in Tao et al. (2013), where the same convergence rate $m^{-\frac{1}{4}} \sqrt{\log n}$ is obtained for a multi-scale realized covariance estimator.

In the empirical implementation, for each pair of assets we choose $H$ as the (rounded) average of the optimal bandwidth for the realized kernel volatility estimator of the two assets following the procedure detailed in Barndorff-Nielsen et al. (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009).

4.2 Factor Structure

Classical asset pricing theory models like the CAPM or APT imply that the unexpected rate of return of risky assets can be expressed as a linear function of few common factors and an idiosyncratic component. Factors induce a fully interconnected partial correlation network structure. In this case, it is natural to carry out network analysis on the partial correlation structure of the assets after netting out the influence of common sources of variation. In this section we propose a modification of our network definition for such systems. Also, we propose a modified covariance estimation strategy analogous to the one put forward in Fan, Fan, and Lv (2008) and Fan et al. (2011) which is based on the particular structure of the system.

We augment the $y$ process with additional $k$ components representing factors. The dynamics of the augmented system are assumed to be the same as the one described in (1). Moreover, the factors are assumed to be observed, as it is commonly done in the empirical finance literature and also as in Fan et al. (2008). The integrated covariance of the augmented system can then
be partitioned as an \((n + k) \times (n + k)\) matrix

\[
\Sigma^* = \begin{bmatrix}
\Sigma_{AA}^* & \Sigma_{FA}^*
\Sigma_{AF}^* & \Sigma_{FF}^*
\end{bmatrix},
\tag{10}
\]

where \(A\) and \(F\) denote, respectively, the blocks of assets and factors.

The covariance of the assets can be expressed as the sum of the systematic and idiosyncratic components, that is

\[
\Sigma_{AA}^* = B \Sigma_{FF}^* B' + \Sigma_I^*,
\]

where

\[
B = \Sigma_{AF}^* [\Sigma_{FF}^*]^{-1} \quad \text{and} \quad \Sigma_I^* = \Sigma_{AA}^* - \Sigma_{AF}^* [\Sigma_{FF}^*]^{-1} \Sigma_{FA}^*.
\]

If the factors are pervasive (\(B\) is not sparse), then the concentration matrix of the assets cannot be sparse. In these cases, rather than proposing a network definition on the basis of the partial correlations of the system, we propose a network definition based on the idiosyncratic partial correlations, that is the partial correlations implied by the idiosyncratic covariance matrix \(\Sigma_I^*\).

Precisely, we define the idiosyncratic integrated partial correlation network as the network whose set of edges is given by

\[
E_I = \{ (i,j) \in V \times V, k_{ij}^* \neq 0, i \neq j \},
\]

where \(k_{ij}^*\) is the \(i,j\)-entry of the matrix \(K_I^* = (\Sigma_I^*)^{-1}\).

Let \(\Sigma\) be an appropriate estimator of the integrated covariance of the augmented system and consider partitioning the estimated covariance matrix analogously to equation (10)

\[
\Sigma = \begin{bmatrix}
\Sigma_{AA} & \Sigma_{FA} \\
\Sigma_{AF} & \Sigma_{FF}
\end{bmatrix}.
\]

Then, a natural estimator of the idiosyncratic realized covariance estimator \(\Sigma_I = (\sigma_{Iij})\) is

\[
\overline{\Sigma}_I = (\overline{\sigma}_{Iij}) = \Sigma_{AA} - \Sigma_{FA} [\Sigma_{FF}]^{-1} \Sigma_{AF}.
\tag{11}
\]

The following corollary establishes the concentration inequality of the estimator \(\overline{\Sigma}_I\) using the one for \(\Sigma\).

**Corollary 1.** If Assumption [A] holds, then there exist positive constants \(b_1, b_2\) and \(b_3\) such that
for all \(i,j \in \{1,\ldots,n\}, x \in [0,b_1]\), and \(M\) large,

\[
P\left( |\sigma_{I_{ij}} - \sigma_{I_{ij}}^*| > x \right) \leq b_2 M^\alpha_0 \exp(-b_3 (M^\beta x)^a),
\]

where \(\beta, a\) and \(\alpha_0\) are the constants from Assumption 7.

The realized network estimator can thus be applied to regularize the idiosyncratic realized covariance matrix and estimate the idiosyncratic partial correlation network. Moreover, the covariance matrix of the assets can be estimated as

\[
\hat{\Sigma}_{AA} = \mathbf{B} \Sigma_{FF} \mathbf{B}' + \hat{\Sigma}_{I\lambda},
\]

where \(\hat{\Sigma}_{I\lambda}\) denotes the realized covariance estimator implied by the realized network. Notice, that this estimation strategy is analogous to the one proposed in Fan et al. (2011).

5 Simulation Study

In this section we carry out a simulation study to assess the finite sample properties of the realized network estimator. The simulation exercise consists of simulating one day of high frequency data and to apply the realized network estimator to estimate the integrated covariance and the integrated concentration matrices. Different specifications of the covariance matrix of the efficient price process are used to assess the usefulness of the realized network estimator depending on the underlying cross-sectional dependence structure of the data. The realized network estimator is also benchmarked against a set of alternative regularization procedures proposed in the literature.

In our simulation study, the efficient price \(y(t)\) is an \(n\)-dimensional zero drift diffusion with \(n\) equal to 50. As in Hautsch et al. (2012), the efficient price follows a simple diffusion with constant covariance \(\Sigma\),

\[
y(t) = \int_0^t \Theta dB(u).
\]

where \(\Theta\) is the Cholesky factorization of \(\Sigma\). The econometrician observes the microstructure contaminated version of the efficient price, and we assume that the noise component \(u_{i,k}\) is an independent zero mean normal random variable with variance equal to \((0.05)^2\) for all stocks. Finally, prices of each stock are observed asynchronously, according to the realization Poisson process with a constant intensity calibrated to have 5 trades per minute on average. In our
numerical implementation, a trading day is 6.5 hours and the simulation of the continuous time process is carried out using the Euler scheme with a discretization step of 5 seconds.

Three different specifications of the covariance matrix $\Sigma$ are adopted. In the first simulation design (Design 1), we pick a specification for $\Sigma$ which induces a sparse partial correlation structure among the assets in the panel. In particular, we choose $\Sigma$ as a function of a realization of an Erdős–Renyi random graph. The Erdős–Renyi random graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph defined over a fixed set of vertices $\mathcal{V} = \{1, \ldots, n\}$ and a random set of edges where the existence of an edge between any pair of vertices is determined by an independent Bernoulli trial with probability $p$. We generate $\Sigma$ by first simulating an Erdős–Renyi random graph $\mathcal{G}$ and then setting $\Sigma$ equal to

$$\Sigma = [I_n + D - A]^{-1},$$

where $I_n$ is the $n$ dimensional identity matrix, and $D$ and $A$ are respectively the degree matrix and the adjacency matrix of the random graph $\mathcal{G}$. The model for $\Sigma$ is such that the underlying random graph structure determines the sparsity structure of the integrated concentration matrix. Also, note that $\Sigma$ is symmetric positive definite by construction. In the simulation we set $p$ equal to $2/n$. In this scenario (i.e. when $np$ is greater than one) the Erdős–Renyi random graph will almost surely have a giant component, that is a connected component containing a constant fraction of the entire set of network vertices. Thus, the highlight of the model is that the generated concentration matrix is sparse while the corresponding covariance matrix is not. In the second simulation design (Design 2), we pick a specification for $\Sigma$ based on a factor model. We set $\Sigma$ as

$$\Sigma = B I_k B' + I_n,$$

where $B$ is an $n \times k$ matrix whose entries are iid normal Gaussian draws with mean zero and unit variance. In the simulation we set the number of factors $k$ to 2. Notice that in this scenario it is challenging for the realized network estimator in that the inverse covariance matrix implied by the model is not sparse. Last, in the third simulation design (Design 3), we set the $\Sigma$ matrix as

$$[\Sigma]_{ij} = \begin{cases} 
\rho + \rho^{|i-j|} & \text{if } i \neq j \\
1 & \text{otherwise}
\end{cases},$$

and set $\rho$ equal to 0.25. Notice that also in this scenario the inverse covariance matrix is not sparse and the covariance matrix does not have a factor representation.
Different approaches are used to estimate the integrated covariance and integrated concentration matrices. First, we estimate the integrated covariance using the 5-minute frequency RC, the pairwise-refreshed time TSRC and the pairwise-refreshed time MRK. The bandwidth parameters of the TSRC and MRK are computed using the plug-in rules previously described. It is important to stress that the TSRC and MRK estimators are not guaranteed to be positive definite. When the estimators are indefinite we apply eigenvalue cleaning (as described in Hautsch et al. (2012)) to obtain a positive definite estimator. For each realized covariance estimator, we consider a number of different regularization procedures. First, we consider the shrinkage estimator proposed by Ledoit and Wolf (2004). Let $\Sigma$ denote the unregularized realized covariance (computed using either the RC, TSRC or MRK estimators and without applying eigenvalue cleaning). The shrinkage estimator is defined as

$$\hat{\Sigma}_{LW} = \alpha_1 I_n + \alpha_2 \Sigma,$$

where $\alpha_1$ and $\alpha_2$ are two tuning parameters chosen to minimize the risk of the estimator that we set following Ledoit and Wolf (2004) and Hautsch et al. (2015). The second regularized estimator is based on a factor approximation of the covariance matrix. It is defined as

$$\hat{\Sigma}_F = \sum_{i=1}^k \hat{\xi}_i \hat{e}_i \hat{e}_i' + \hat{R}_k,$$

where $\hat{\xi}_i$ and $\hat{e}_i$ denote the eigenvalues (in increasing order) and corresponding eigenvectors obtained from the spectral decomposition of the unregularized realized estimator $\Sigma$, and $\hat{R}_k$ is diag($\Sigma - \sum_{i=1}^k \hat{\xi}_i \hat{e}_i \hat{e}_i'$). Notice that the shrinkage and factor regularization procedures are also not guaranteed to provide a positive definite estimator when applied to the TSRC and MRK estimators. In these cases we apply eigenvalue cleaning whenever the resulting estimator is not positive definite (which however happens rarely in the simulation designs we consider). Last, we use the realized network estimator defined in Section 2 using the BIC criterion to determine the optimal amount of shrinkage to apply. Notice that one of the inputs of the BIC criterion is the number of observations used to compute the estimator. When using pairwise-refresh sampling however, this number is different for each entry of the covariance matrix. Similarly to Hautsch et al. (2012) we opt for a conservative choice of this quantity, and we set it to the minimum number of refresh time observations across all pairs. The realized network estimator
is guaranteed to be positive definite as long as the pre-estimator $\Sigma$ is positive semi-definite. If the unregularized estimator is indefinite, we add to it the identity matrix times the absolute value of its smallest eigenvalue to obtain a positive semi-definite estimator.

Different metrics are used to evaluate the performance of the estimators. A classic loss function used for the evaluation of covariance matrix estimators is the Kullback–Liebler loss proposed by Stein ([Stein 1956](#) [Pourahmadi 2013](#)), which is defined as

$$\text{KL}(\hat{\Sigma}) = \text{tr}(\hat{\Sigma}K^*) - \log |\hat{\Sigma}K^*| - n.$$ 

Following ([Hautsch et al. 2012](#)), we also consider a Root Means Square Error (RMSE) type loss based on the scaled Frobenius norm of the covariance matrix, which is defined as

$$\text{RMSE}(\hat{\Sigma}) = \sqrt{\frac{1}{n} \sum_{i,j=1}^{n} (\hat{\sigma}_{ij} - \sigma^*_ij)^2}.$$ 

We perform 10’000 Monte Carlo replications of the simulation exercise for each simulation design and report summary statistics on the performance of the estimators in table [1](#). The table reports the average of the KL and RMSE losses of the estimators in the three simulation designs. Results convey that using a regularization technique whose shrinking target is closer to the true underlying structure of the data produces the best results and large improvements over the unregularized realized covariance matrix estimator. In particular, it is easy to see that when the partial correlation structure of the data is sparse (Design 1) the realized network estimator is the best performing regularization technique. Analogously, the factor based regularization works best in Design 2 and shrinkage regularization works in Design 3. Overall, the results convey that the gains by using the realized network estimator when the partial correlation structure is sparse can be substantial.

INSERT TABLE [1](#) ABOUT HERE

6 Empirical Application

We use the realized network estimator to analyse the dependence structure of a panel of US bluechips from the NYSE throughout 2009. We then engage in a Markowitz style Global Minimum Variance portfolio prediction exercise to highlight the advantages of the methodology for forecasting.
6.1 Data and Estimation

We consider a sample of 100 liquid US bluechips that have been part of the S&P100 index for most of the 2000’s. We also include in the panel the SPY ETF, the ETF tracking the S&P 500 index. We work with tick-by-tick transaction prices obtained from the NYSE-TAQ database. Before proceeding with the econometric analysis, the data are filtered using standard techniques described in Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). The full list of tickers, company names and industry groups is reported in table 2.

We estimate the integrated covariance of the assets throughout 2009. More precisely, for each of the 52 weeks of 2009, we use the data on the last weekday available of each week to construct the realized covariance estimators. On each of these days, we only consider the tickers that have at least 1000 trades. Exploratory analysis (not reported in the paper) confirms the well documented evidence of a CAPM-type factor structure in the panel. To this extent, our realized covariance estimation strategy consists of first decomposing the realized covariance in systematic and idiosyncratic covariance components and then regularizing the idiosyncratic part with the realized network. More precisely, we compute the realized covariance of the assets in the panel together with the SPY ticker (the proxy of the market), and then obtain the systematic and idiosyncratic components of the realized covariance of the assets on the basis of formula (11). Finally, we apply the realized network regularization procedure to the idiosyncratic realized covariance. On each week of 2009, we estimate the realized network using three (idiosyncratic) realized covariance estimators: MRK, TSRC as well as the classic RC using data sampled at a 1 minute frequency.

6.2 Realized Network Estimates

In this section we present the realized network estimation results. We first provide details for one specific date only that roughly corresponds to the mid of the sample (June 26, 2009), and then report summaries for all estimated networks in 2009. In the interest of space we report the TSRC estimator results only. The MRK and RC provide similar evidence.

We begin by showing in figure 1 the heatmap of the idiosyncratic correlation matrix associated with the idiosyncratic realized covariance estimator on June 26. Notice that the heatmap
is constructed by sorting stocks by industry group and then by alphabetical order. The picture clearly shows that after netting out the influence of the market factor, a fair amount of cross-sectional dependence is still present across stocks. Inspection of the heatmap reveals that the majority of estimated correlation coefficients are positive. The correlation matrix exhibits a block diagonal structure hinting that correlation is stronger among firms in the same industry. On this date, the intra-industry group correlation is particularly strong for energy companies.

We estimate the realized network using the glasso and use the bic to choose the optimal amount of shrinkage. To give insights on the sensitivity of the estimated network to the shrinkage parameter $\lambda$, the top panel of figure 2 reports the so called trace, which is the graph of the estimated partial correlations as a function of the shrinkage parameter. The bottom panel of the same figure also shows the value of the bic criterion as a function of the shrinkage parameter. The plot highlights how the sparsity of the estimated network varies substantially over the range of lambda values considered and that the majority of estimated partial correlations are positive irrespective of the value of shrinkage imposed on the estimator. The estimated network corresponding to the optimal $\lambda$ has 188 edges, which correspond to approximately the 5% of the total number of possible edges in the network on this date. The number of companies that have interconnections are 66 (roughly 2/3 of the total) and are all connected in a unique giant component.

It is useful to provide details on the amount of variability explained by the systematic and idiosyncratic components of the covariance matrix of the panel. To this extent, we introduce the systematic coefficient of determination, defined as

$$R^2_{Fi} = \frac{B_i^\prime \Sigma_{FF} B_i}{B_i^\prime \Sigma_{FF} B_i + \hat{\sigma}_{Iii}},$$

which measures the amount of variability of asset $i$ explained by the market factor. We also introduce the idiosyncratic coefficient of determination, defined as

$$R^2_{Ii} = \frac{\hat{\sigma}_{Iii} - 1/\hat{k}_{Iii}}{\hat{\sigma}_{Iii}},$$

where $\hat{k}_{Iii}$ is the estimated number of factors.
which measures the amount of variability of asset $i$ explained by the remaining assets conditional on the market factor. On June 26, the average of the systematic $R^2_F$ is equal to 22.8% while the average of the idiosyncratic $R^2_I$ (for those assets with at least one neighbour) is 7.3%. Overall the systematic component is the most relevant one in terms of explained variability, however, the idiosyncratic component captures a non negligible portion of variability as well.

Figure 3 displays the idiosyncratic partial correlation network. A number of comments on the empirical characteristics of the network are in order. First, on this date, Google (GOOG) emerges as a particularly highly interconnected firm, with linkages spreading to several other industry groups. The estimated network also shows some degree of industry clustering, that is linkages are more frequent among firms in the same industry group. In order to get better insights on the industry linkages in table 3 we report the total number of links across industry groups. The table shows that firms in the industrials, energy, technology and financials groups are particularly interconnected among each other. On the other hand, consumer discretionary, cosumer staples and healthcare have few intra-industry linkages. In figure 4 we report the degree distribution of the estimated network and the distribution of the nonzero partial correlations. As far as the degree distribution is concerned, the network exhibits the typical features of Power Law networks, that is the number of connections is heterogeneous and the most interconnected stocks have a large number of connections relative to the total number of links. The histogram of the partial correlation shows that the majority of the partial correlations are positive and that positive partial correlations are on average larger than the negative ones.

Last, we are interested in determining which companies are more interconnected and central in the network. We measure the degree of interconnectedness of a firm using different approaches: (i) the degree of a company in the network (that is, the number of links); (ii) the sum of nonzero square partial correlations of a company; and (iii) the centrality index of the page–rank algorithm. The page–rank algorithm is a famous network based centrality index used by web search engines to rank web pages. It turns out that the indices provide substantially close rankings. The rank correlation among the different measures are all above 0.9. We report the top ten most central companies in table 4 according to page rank. The page rank algorithm shows that Google is indeed the most central stock on this date.

We report a number of summary statistics for the sequence of networks estimated in 2009.
First, in figure 5 we report the proportion of number of links in the network throughout the year. The picture shows that sparsity is stable at a value slightly below 5% of the total possible number of linkages. The plots show the sparsity rate vis-à-vis the VIX volatility index to show that no particular time series pattern emerges in the network sparsity and that in particular the sparsity is unrelated to the level of volatility of the market.

Figure 6 shows the total number of links of each industry group divided by the total number of possible edges. The plot omits the series for materials, telecom and utilities due to their small size. Technology, energy, financial and industrials are the most interconnected sectors also throughout 2009, and the level and cross sectional rankings are fairly stable across the sample. In order to give more insights on the degree of concentration within each group, in figure 7 we report the concentration of links in each industry group measured using the Herfindahl index. Again, materials, telecom and utilities are omitted from the graph. Once again, no particular time series pattern emerges from the plot and cross sectional concentration rankings are quite stable. The picture shows that the most interlinked sectors have quite different concentration characteristics. Technology is one of the most highly concentrated sector. Detailed inspection of the results reveals that this is driven by the fact that in 2009 Google is essentially the most interconnected ticker in the sample. On the other hand, industrials have the smaller average concentration, in that the number of links is quite uniformly distributed across firms and no specific “hub” emerges among these tickers.

Overall results convey that after conditioning on a one factor structure that data still has a fair amount of cross-sectional dependence and that networks provide a useful device to synthesize such information. The main empirical features of the network are stable throughout 2009. Firms in the energy and industrials sectors are strongly interconnected. Technology companies and Google in particular are the most highly interconnected firms throughout the year. We conjecture that this could be driven by the fact that starting from March 2009, with the beginning of post crisis market rally, technology stocks and Google in particular have been the most rapidly recovering stocks of the year.
6.3 Predictive Analysis

In order to assess the ability of the regularized network methodology to provide more precise estimates of the integrated covariance we carry out an asset allocation prediction exercise \cite{Hautsch2012, Hautsch2015, EngleColacito2006}. The forecasting exercise is designed as follows. For each week of 2009, we construct the Markowitz Global Minimum Variance (GMV) portfolio weights using the formula

\[ \hat{w} = \hat{\Sigma}^{-1} 1 
\]

where \(1\) is \(n\)-dimensional vector of ones. The resulting GMV portfolio weights are used to construct a portfolio of the assets which is held for the following week. At the end of the week, we compute the daily sample variance of the portfolio return for that period. We repeat the exercise for all the weeks of 2009. The performance of the covariance estimators is evaluated by assessing which estimator delivers the smallest average out-of-sample GMV portfolio variance.

The set of estimators we consider are based on the systematic/idiosyncratic decomposition of the covariance matrix

\[ \hat{\Sigma} = \mathbb{B} \Sigma_F \mathbb{B}' + \hat{\Sigma}_I, \]

and differ on the choice of the estimator of the idiosyncratic realized covariance matrix \(\hat{\Sigma}_I\). The set of candidate idiosyncratic realized covariance estimators contains: (i) unregularized covariance estimator; (ii) constrained covariance estimator, obtained by setting all the off-diagonal elements of the unregularized covariance estimator to zero; (iii) shrinkage covariance estimator of \cite{LedoitWolf2004} (see equation (12)); (iv) factor regularized covariance estimator (see equation (13)) based on three factors; (v) block-factor regularized estimator, obtained by applying factor regularization of equation (13) based on one factor to each industry block and setting the rest of the covariance matrix to zero; and (vi) realized network estimator. The exercise is carried out using the MRK, TSRC and RC estimators.

We report summary results on the forecasting exercise in table 5. The table shows the average annualized volatility of the GMV portfolios. The three different covariance estimators deliver analogous inference. The constrained estimator that ignores cross-sectional dependence in the idiosyncratic realized covariance matrix typically performs worst than the baseline unconstrained realized covariance estimators. Interestingly, the factor and block factor regularization schemes do not produce large gains out-of-sample in comparison to the benchmark. We interpret this as the consequence that after controlling for the market factor there is only weak evidence of
the presence of additional factors. The block-factor regularization might be not particularly successful because while some sectors exhibit strong dependence (Industrials) a large number of stocks in other sectors do not (Consumer Discretionary, Consumer Staples and Healthcare). The shrinkage and realized network regularization schemes provide the best out–of–sample results, and the realized network estimator in particular achieves the lower out–of–sample variance. Also, it is interesting to point out that among the two market friction robust estimators, the MRK delivers lower out of sample losses than the TSRC. Last we note that the difference in the forecasting performance across the different realized volatility estimators is substantially smaller after carrying out network regularization.

7 Conclusions

In this work we propose a regularization procedure for realized covariance estimators. The regularization consists of shrinking the off–diagonal elements of the inverse realized covariance matrix to zero using a Lasso–type penalty. Since estimating a sparse inverse realized covariance matrix is equivalent to detecting the partial correlation network structure of the daily log-prices, we call our procedure realized network. The technique is specifically designed for the Two–Scales Realized Covariance (TSRC) and the Multivariate Realized Kernel (MRK) estimators based on refresh time sampling, which are state–of–the–art consistent covariance estimators that allow for market microstructure effects and asynchronous trading. We establish the large sample properties of the procedure estimator and show that the realized network consistently estimates the inverse integrated covariance matrix and consistently detects the nonzero partial correlations of the network. An empirical exercise is used to highlight the usefulness of the procedure and an out–of–sample GMV portfolio asset allocation exercise is carried out to compare our procedure against a number of benchmarks. Results convey that realized network enhances the prediction properties of classic realized covariance estimators and performs well relative to a set of alternative regularization procedures.
References


Table 1: Simulation Study

<table>
<thead>
<tr>
<th></th>
<th>No regular.</th>
<th>Shrinkage Factor Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KL</td>
<td>65.89</td>
<td>55.13</td>
</tr>
<tr>
<td>RMSE</td>
<td>55.77</td>
<td>44.98</td>
</tr>
<tr>
<td>TSRC</td>
<td>51.95</td>
<td>48.72</td>
</tr>
<tr>
<td>KL</td>
<td>29.97</td>
<td>29.56</td>
</tr>
<tr>
<td>RMSE</td>
<td>53.44</td>
<td>48.64</td>
</tr>
<tr>
<td>MRK</td>
<td>32.50</td>
<td>29.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>73.54</td>
<td>52.07</td>
</tr>
<tr>
<td>RMSE</td>
<td>68.65</td>
<td>54.59</td>
</tr>
<tr>
<td>TSRC</td>
<td>52.30</td>
<td>10.99</td>
</tr>
<tr>
<td>RMSE</td>
<td>30.09</td>
<td>20.84</td>
</tr>
<tr>
<td>MRK</td>
<td>60.67</td>
<td>46.32</td>
</tr>
<tr>
<td>RMSE</td>
<td>31.03</td>
<td>22.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>36.29</td>
<td>15.53</td>
</tr>
<tr>
<td>RMSE</td>
<td>39.46</td>
<td>19.28</td>
</tr>
<tr>
<td>TSRC</td>
<td>40.26</td>
<td>4.64</td>
</tr>
<tr>
<td>RMSE</td>
<td>15.95</td>
<td>10.65</td>
</tr>
<tr>
<td>MRK</td>
<td>28.34</td>
<td>4.59</td>
</tr>
<tr>
<td>RMSE</td>
<td>19.38</td>
<td>10.52</td>
</tr>
</tbody>
</table>

The table reports the KL and RMSE average losses of the unregularized RC, TSRC, and MRK estimators (No regular) as well as their regularized versions (Shrinkage, Factor, Network) in the three simulation designs of Section 5.
<table>
<thead>
<tr>
<th>Ticker Symbol</th>
<th>Company Name</th>
<th>GICS Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMZN</td>
<td>Amazon.com</td>
<td>Health Care</td>
</tr>
<tr>
<td>ABT</td>
<td>Abbott Laboratories</td>
<td>Health Care</td>
</tr>
<tr>
<td>CMCSA</td>
<td>Comcast</td>
<td>Consumer Discretionary</td>
</tr>
<tr>
<td>DIS</td>
<td>Walt Disney</td>
<td>Consumer Discretionary</td>
</tr>
<tr>
<td>BAX</td>
<td>Baxter International</td>
<td>Health Care</td>
</tr>
<tr>
<td>MDK</td>
<td>Southern Company For Energy</td>
<td>Health Care</td>
</tr>
<tr>
<td>BA</td>
<td>Boeing Company</td>
<td>Industrials</td>
</tr>
<tr>
<td>TWX</td>
<td>Time Warner Inc.</td>
<td>Consumer Discretionary</td>
</tr>
<tr>
<td>COST</td>
<td>The Coca-Cola Company</td>
<td>Consumer Staples</td>
</tr>
<tr>
<td>GOOGL</td>
<td>Alphabet Inc.</td>
<td>Information Technology</td>
</tr>
<tr>
<td>AAPL</td>
<td>Apple Inc.</td>
<td>Information Technology</td>
</tr>
<tr>
<td>XOM</td>
<td>Exxon Mobil</td>
<td>Energy</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machines</td>
<td>Information Technology</td>
</tr>
<tr>
<td>BAC</td>
<td>Bank of America</td>
<td>Financials</td>
</tr>
<tr>
<td>MSFT</td>
<td>Microsoft</td>
<td>Information Technology</td>
</tr>
<tr>
<td>JNJ</td>
<td>Johnson &amp; Johnson</td>
<td>Health Care</td>
</tr>
<tr>
<td>JPM</td>
<td>JPMorgan Chase &amp; Co.</td>
<td>Financials</td>
</tr>
<tr>
<td>TGT</td>
<td>Target</td>
<td>Consumer Discretionary</td>
</tr>
<tr>
<td>UNP</td>
<td>Union Pacific</td>
<td>Industrials</td>
</tr>
<tr>
<td>MRK</td>
<td>Merck</td>
<td>Health Care</td>
</tr>
<tr>
<td>MRCK</td>
<td>Merck &amp; Co.</td>
<td>Health Care</td>
</tr>
<tr>
<td>PFE</td>
<td>Pfizer</td>
<td>Health Care</td>
</tr>
<tr>
<td>SBUX</td>
<td>Starbucks</td>
<td>Consumer Discretionary</td>
</tr>
<tr>
<td>VZ</td>
<td>Verizon</td>
<td>Telecommunications</td>
</tr>
<tr>
<td>V</td>
<td>Verizon Communications</td>
<td>Telecommunications</td>
</tr>
<tr>
<td>SNW</td>
<td>Southern Company For Energy</td>
<td>Health Care</td>
</tr>
</tbody>
</table>

The table reports the list of company tickers, company names and industry sectors.
Table 3: Links on 2009-06-26

<table>
<thead>
<tr>
<th>Disc</th>
<th>Stap</th>
<th>Ener</th>
<th>Fin</th>
<th>Heal</th>
<th>Ind</th>
<th>Tech</th>
<th>Mat</th>
<th>Tel</th>
<th>Util</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>4</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stap</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ener</td>
<td>1</td>
<td></td>
<td>35</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fin</td>
<td></td>
<td>11</td>
<td></td>
<td>2</td>
<td>14</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heal</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>20</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech</td>
<td>8</td>
<td>6</td>
<td>12</td>
<td>14</td>
<td>7</td>
<td>20</td>
<td>14</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Mat</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Util</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

The table reports the number of estimated links among industry groups on June 26, 2009.

Table 4: Centrality on 2009-06-26

<table>
<thead>
<tr>
<th>Rank</th>
<th>Ticker</th>
<th>Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GOOG</td>
<td>Information Technology</td>
</tr>
<tr>
<td>2</td>
<td>MA</td>
<td>Information Technology</td>
</tr>
<tr>
<td>3</td>
<td>SLB</td>
<td>Energy</td>
</tr>
<tr>
<td>4</td>
<td>FCX</td>
<td>Materials</td>
</tr>
<tr>
<td>5</td>
<td>APA</td>
<td>Energy</td>
</tr>
<tr>
<td>6</td>
<td>COP</td>
<td>Energy</td>
</tr>
<tr>
<td>7</td>
<td>OXY</td>
<td>Energy</td>
</tr>
<tr>
<td>8</td>
<td>APC</td>
<td>Energy</td>
</tr>
<tr>
<td>9</td>
<td>DVN</td>
<td>Energy</td>
</tr>
<tr>
<td>10</td>
<td>CVX</td>
<td>Energy</td>
</tr>
</tbody>
</table>

The table reports the top tickers by eigenvector centrality on June 26, 2009.

Table 5: GMV Forecasting

<table>
<thead>
<tr>
<th></th>
<th>No regular</th>
<th>Diagonal</th>
<th>Network</th>
<th>Shrinkage</th>
<th>Factor</th>
<th>Block-Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>39.10</td>
<td>40.53</td>
<td>26.16</td>
<td>31.86</td>
<td>31.38</td>
<td>32.68</td>
</tr>
<tr>
<td>TSRRC</td>
<td>37.22</td>
<td>41.41</td>
<td>26.22</td>
<td>29.38</td>
<td>30.60</td>
<td>31.58</td>
</tr>
<tr>
<td>MRK</td>
<td>32.83</td>
<td>35.51</td>
<td>24.52</td>
<td>27.81</td>
<td>28.49</td>
<td>28.02</td>
</tr>
</tbody>
</table>

The table reports the results of the GMV forecasting comparison exercise. The table reports the annualized out of sample volatilities of the GMV portfolios constructed for the unregularized RC, TSRC and MRK estimators (No regular.) as well as their regularized versions (Diagonal, Network, Shrinkage, Factor, Block-Factor).
Figure 1: **Idiosyncratic Correlation Heatmap on 2009-06-26**

The figure shows the heatmap of the idiosyncratic realized correlation matrix on June 26, 2009 estimated using the TSRC estimator. Darker colors indicate higher correlations in absolute value.

Figure 2: **Trace on 2009-06-26**

The figure shows the trace and BIC of the realized network estimator on June 26, 2009.

32
Figure 3: Idiosyncratic Partial Correlation Network on 2009-06-26

The figure shows the optimal realized network estimated on June 26, 2009.

Figure 4: Degree and Partial Correlation Distribution on 2009-06-26

The figure shows the degree distribution and the distribution of partial correlations on June 26, 2009.
Figure 5: **Sparsity vs. Volatility**

The figure shows the sparsity of the estimated network (square) vis-à-vis the level of volatility measured by the VIX (circle) for each week of 2009.

Figure 6: **Sectorial Links**

The figure shows the number of linkages of the different industry groups over the total number of possible linkages for each week of 2009.
Figure 7: Sectorial Concentration

The figure shows the link concentration (measured using the Herfindahl index) of the different industry for each week of 2009.
A Technical Appendix

Following the same notation as in Ravikumar et al. (2011), we set
\[
\tilde{K} = \arg \min_{K \in S^n, K_{sc} = 0} \left\{ \text{tr}(\Sigma K) - \log \det(K) + \lambda \sum_{i \neq j} |k_{ij}| \right\},
\]

\[W = (w_{ij}) = \Sigma - \Sigma^*, \triangle = \tilde{K} - K^*, R(\triangle) = \tilde{K} - K^* + K^* \triangle K^*.
\]

We need the following lemma in order to prove Theorem 1 for \(\alpha_0 = 1\).

**Lemma 1.** Assume that \(m \geq m_0 := \left( -\frac{2}{a_3} \log \left( \frac{a_1}{a_2} x^\frac{1}{\beta} \right)^{\frac{1}{\beta}} \right) \), for some \(\tau > 2\), where \(a_i, a\) and \(\beta\) are the constants in Assumption 1. Then for all \(x \in \left[0, \min \left( a_1, \frac{a}{a_3} \exp \left( -\frac{2}{a^2 \beta} \right) \right) \right]\),

\[P (\|W\|_\infty \geq x) \leq \frac{1}{n^{\tau-2}}.
\]

**Proof.** According to Assumption 1 when \(x \in [0, a_1]\), for generic \(i, j\) we have

\[P (|w_{ij}| \geq x) \leq a_2 m \exp \left( -a_3 \left( m^\beta x \right)^a \right) = P_1 P_2,
\]

where \(P_1 = \frac{a_2}{a_3} x^{-\frac{1}{\beta}} \exp \left( -a_3 \left( m^\beta x \right)^a \right), P_2 = \left( a_3 m a^\beta x^a \right)^{\frac{1}{\beta}} \exp \left( -a_3 \left( m^\beta x \right)^a \right)\). \(P_1\) is decreasing in \(m\). Thus, when \(m \geq m_0, P_1 \leq n^{-\tau}\). On the other hand, \(a_3 m a^\beta x^a \geq -2 \log \left( \frac{a_1}{a_2} x^\frac{1}{\beta} \right)^{\frac{1}{\beta}}\).

Thus, when \(x \leq \frac{a^\beta}{a_3} \exp \left( -\frac{2}{a^2 \beta} \right), a_3 m a^\beta x^a \geq -2 \log \left( \frac{a_1}{a_2} x^\frac{1}{\beta} \right) \geq \frac{1}{a^{2 \beta}}\).

Consider the function \(f(y) = \left( y \exp \left( -\frac{a^\beta y}{2} \right) \right)^{\frac{1}{\beta}}\). It is easy to see that when \(y \geq \frac{4}{a^2 \beta^2}\), \(f(y) \leq 1\). As \(P_2 = f \left( a_3 m a^\beta x^a \right), a_3 m a^\beta x^a \geq \frac{4}{a^2 \beta^2}\), we get that \(P_2 \leq 1\). Since \(P_1 \leq n^{-\tau}\), we obtain that

\[P (|w_{ij}| \geq x) \leq n^{-\tau}.
\]

Using this inequality over all \(n^2\) entries of \(W\), we conclude that

\[P (\|W\|_\infty \geq x) = P \left( \max_{i,j} |w_{ij}| \geq x \right) \leq n^{2-\tau}.
\]

**Proof of Theorem 1.** First we prove (a). Set \(h(x) = a_2 m \exp \left( -a_3 \left( m^\beta x \right)^a \right)\) for \(x > 0\), and
\[ a_0 := \min \left( a_1, \frac{a_{\beta}}{a_3} \exp \left( -\frac{2}{a_{\beta}} \right) \right). \]  
In the proof of Lemma 1, we have shown that when \( m \geq m_0 \), which is true according to (5) and (6), \( h(a_0) \leq n^{-\tau} \). On the other hand, \( h \left( \frac{a}{\alpha} \lambda \right) = n^{-\tau} \). Because the function \( h \) is decreasing, we get that \( \frac{a}{\alpha} \lambda \leq a_0 \). Therefore, by Lemma 1, we conclude that

\[
P \left( \|W\|_\infty \leq \frac{\alpha}{8} \lambda \right) \geq 1 - \frac{1}{n^{\tau-\tau}}. \tag{15}\]

By the same argument, if \( \sigma_m \leq a_0 \), as \( m \geq \left( -\frac{2}{a_3 a_{\beta}} \log \left( \frac{\alpha}{a_2 - \sigma_m n^{-\tau}} \right) \right)^{\frac{1}{a_3}}, \) we have \( h(\sigma_m) \leq n^{-\tau}, \) so \( \frac{\alpha \lambda}{8} \leq \sigma_m \). When \( \sigma_m > a_0 \), we still have \( \frac{\alpha \lambda}{8} \leq \sigma_m \), since \( \frac{\alpha \lambda}{8} \leq a_0 \). Therefore,

\[
P \left( \|W\|_\infty \leq \sigma_m \right) \geq P \left( \|W\|_\infty \leq \frac{\alpha}{8} \lambda \right) \geq 1 - \frac{1}{n^{\tau-\tau}}. \tag{16}\]

and when the event \( A := \left\{ \|W\|_\infty \leq \frac{\alpha}{8} \lambda = m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}} \right\} \) holds, \( \Sigma \) has positive diagonal entries, and by Lemma 3 in Ravikumar et al. (2011), \( \tilde{K} \) is unique. Proceeding as before, we can obtain that \( \left( 6 \left( 1 + 8\alpha^{-1} \right)^2 d \max \left( C_{\Sigma}, C_{\Gamma}, C_{\tilde{\Sigma}}, C_{\tilde{\Gamma}} \right)^{-1} \right) \leq \frac{\alpha \lambda}{8} \). Thus, we get that, if \( A \) holds, then

\[
2C_{\Gamma} \left( \|W\|_\infty + \lambda \right) \leq 2C_{\Gamma} \left( 1 + 8\alpha^{-1} \right) m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}} \leq \left( 3d \max \left( C_{\Sigma}, C_{\tilde{\Sigma}}, C_{\tilde{\Gamma}} \right)^{-1} \right)^{-1}. \]

Therefore, by Lemma 6 in Ravikumar et al. (2011), it holds that

\[
\|\Delta\|_\infty \leq 2C_{\Gamma} \left( 1 + 8\alpha^{-1} \right) m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}}. \]

Now, appealing to Lemma 5 in Ravikumar et al. (2011), we obtain that

\[
\|R(\Delta)\|_\infty \leq \frac{3}{2} d \|\Delta\|_\infty C_{\tilde{\Sigma}}^3 C_{\tilde{\Gamma}} \leq 6C_{\tilde{\Sigma}}^3 C_{\tilde{\Gamma}} d \left( 1 + 8\alpha^{-1} \right)^2 \left( m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}} \right)^2 \]

\[
= 6C_{\tilde{\Sigma}}^2 C_{\tilde{\Gamma}} d \left( 1 + 8\alpha^{-1} \right)^2 m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}} \leq \frac{\alpha \lambda}{8} \leq \frac{\alpha \lambda}{8}. \]

By Lemma 4 in Ravikumar et al. (2011), we conclude that \( \tilde{K} = \hat{K} \). Thus, \( \|\hat{K} - K^*\|_\infty \) satisfies the same bound as \( \|\Delta\|_\infty \).

Now let us prove (b). Using the same argument as before, it is easy to see that

\[
\frac{k_m}{2C_{\Gamma} \left( 1 + 8\alpha^{-1} \right)} \geq \frac{\alpha \lambda}{8} = m^{-\beta} \left( \frac{\log(a_{2mn^r})}{a_3} \right)^{\frac{1}{a_3}}. \]
Then we have
\[ k_m \geq 2C_T \cdot (1 + 8\alpha^{-1}) m^{-\beta} \left( \frac{\log(a_{2mn}^T)}{a_3} \right)^{\frac{1}{m}} \geq 2C_T \cdot (||W||_\infty + \lambda). \]

Finally appealing to Lemma 7 in Ravikumar et al. (2011), the proof is completed. \( \square \)

**Proof of Corollary 7.** Notice that \( \Sigma_{FF}^* = \sigma_{n+1,n+1}^* \) and \( \Sigma_{FF} = \sigma_{n+1,n+1} \). Then for all \( i, j \in \mathcal{V} \),
\[ \sigma_{ij}^* = \sigma_{ij} - \frac{\sigma_{n+1,i}^* \sigma_{j,n+1}^*}{\sigma_{n+1,n+1}^*} \quad \text{and} \quad \sigma_{ij} = \sigma_{ij} - \frac{\sigma_{n+1,i} \sigma_{j,n+1}}{\sigma_{n+1,n+1}}. \]

Therefore,
\[ P( |\sigma_{ij} - \sigma_{ij}^*| > x) \leq P \left( \left| \sigma_{ij}^* - \sigma_{ij} \right| \geq \frac{x}{2} \right) + P \left( \left| \frac{\sigma_{n+1,i} \sigma_{j,n+1}}{\sigma_{n+1,n+1}} - \frac{\sigma_{n+1,i}^* \sigma_{j,n+1}^*}{\sigma_{n+1,n+1}^*} \right| \geq \frac{x}{2} \right). \]

To the first term, we can apply the concentration inequality \( \lbrack 2 \rbrack \). In order to bound the second term, set \( b_1 = \sigma_{n+1,n+1} - \sigma_{n+1,n+1}^* \), \( b_2 = \sigma_{n+1,i} - \sigma_{n+1,i}^* \), and \( b_3 = \sigma_{j,n+1} - \sigma_{j,n+1}^* \). Then,
\[ \frac{\sigma_{n+1,i} \sigma_{j,n+1}}{\sigma_{n+1,n+1}} - \frac{\sigma_{n+1,i}^* \sigma_{j,n+1}^*}{\sigma_{n+1,n+1}^*} = \left| \frac{b_3 \sigma_{n+1,n+1}^* \sigma_{n+1,i}^* + b_2 \sigma_{n+1,n+1}^* \sigma_{j,n+1}^* + b_2 b_3 \sigma_{n+1,n+1}^* - b_1 \sigma_{n+1,i}^* \sigma_{j,n+1}^*}{\sigma_{n+1,n+1}^* + b_1} \right|. \]

Without loss of generality, we assume that \( \sigma_{j,n+1}^* \) and \( \sigma_{n+1,i}^* \) are non-zero. Otherwise, the proof follows similarly. Then,
\[ P \left( \left| \frac{\sigma_{n+1,i} \sigma_{j,n+1}}{\sigma_{n+1,n+1}} - \frac{\sigma_{n+1,i}^* \sigma_{j,n+1}^*}{\sigma_{n+1,n+1}^*} \right| \geq 8x \right) \leq P \left( |b_1| \geq \min \left( \frac{\sigma_{n+1,n+1}^*}{2}, \frac{\sigma_{n+1,n+1}^*}{|\sigma_{j,n+1}^*|} \right) \right) + P \left( |b_2| \geq \min \left( \frac{\sigma_{n+1,n+1}^*}{|\sigma_{j,n+1}^*|}, 1 \right) \right) + P \left( |b_3| \geq \frac{\sigma_{n+1,n+1}^*}{|\sigma_{j,n+1}^*|} \right). \]

Observe that in the last inequality we have used the fact that if \( |b_1| < \frac{\sigma_{n+1,n+1}^*}{2} \), then we can lower bound the denominator in \( \lbrack 17 \rbrack \) by \( \frac{\sigma_{n+1,n+1}^*}{2} \). Then appealing again to the concentration inequality \( \lbrack 2 \rbrack \), we conclude the desired result. \( \square \)

**Proof of Theorem 3.** We start proving the concentration inequality for a generic asset \( i \). To simplify the exposition, we assume that the drift is zero in \( \lbrack 1 \rbrack \), otherwise it is easy to see that
the same result follows. By definition (9), the realized kernel estimator of \( \sigma_{ii}^* \) equals

\[
\pi_{Kii} = \gamma_0(y, y) + A + B + C,
\]

where

\[
A = \sum_{h=1}^{H} k \left( \frac{h - 1}{H} \right) (\gamma_h(y, y) + \gamma_{-h}(y, y)),
\]

\[
B = 2\gamma_0(y, u) + \sum_{h=1}^{H} k \left( \frac{h - 1}{H} \right) (\gamma_h(y, u) + \gamma_h(u, y) + \gamma_{-h}(y, u) + \gamma_{-h}(u, y)),
\]

\[
C = \gamma_0(u, u) + \sum_{h=1}^{H} k \left( \frac{h - 1}{H} \right) (\gamma_h(u, u) + \gamma_{-h}(u, u)),
\]

and for each \( h \in \{-H, \ldots, H\} \), \( \gamma_h(y, u) = \sum_{j=1}^{m-1} (y_j - y_{j-1})(u_{j-h} - u_{j-h-1}) \).

By Lemma 3 in Fan et al. (2012), there exist constants \( c_1, c_2 > 0 \) such that for large \( m \) and \( x \in [0, c_1] \),

\[
P \left( |\gamma_0(y, y) - \sigma_{ii}^*| \geq x \right) \leq 2 \exp(-c_2mx^2).
\]

Therefore, by (18) and (19), it suffices to bound the terms \( A, B \) and \( C \) in probability.

Bound in probability of the term \( A \) in (18). We decompose \( A \) as

\[
A = \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u)dB(u) \sum_{h=1}^{H} k \left( \frac{h - 1}{H} \right) \left( \int_{t_{j-h-1}}^{t_{j-h}} \theta(u)dB(u) + \int_{t_{j+h-1}}^{t_{j+h}} \theta(u)dB(u) \right)
\]

\[
= A_1 + A_2.
\]

We start bounding \( A_1 \). For any \( x > 0 \) and \( \alpha > 0 \), we have

\[
P \left( |A_1| \geq \frac{x}{2} \right) \leq P \left( \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u)dB(u)D_j \mathbf{1} \left\{ \sup_{j \in \{1, \ldots, m-1\}} |D_j| \leq m^{-\frac{1}{4} + \alpha} \right\} \geq \frac{x}{4} \right)
\]

\[
+ P \left( \sup_{j \in \{1, \ldots, m-1\}} |D_j| > m^{-\frac{1}{4} + \alpha} \right),
\]

where

\[
D_j = \sum_{h=1}^{H} k \left( \frac{h - 1}{H} \right) \int_{t_{j-h-1}}^{t_{j-h}} \theta(u)dB(u).
\]
Since \( k \) and \( \theta \) are bounded, it holds that
\[
\sum_{h=1}^{H} k^2 \left( \frac{h - 1}{H} \right) \int_{t_{j-h-1}}^{t_{j-h}} |\theta(u)|^2 \, du \leq c H \frac{C_\triangle}{m} = cm^{-\frac{1}{2}}.
\]
Therefore, appealing to the exponential martingale inequality, we get that
\[
P \left( \sup_{j \in \{1, \ldots, m-1\}} |D_j| > m^{-\frac{1}{2} + \alpha} \right) \leq 2m \exp \left( -cm^{2\alpha} \right).
\]
On the other hand, since
\[
\int_{t_{j-1}}^{t_j} |\theta(u)|^2 \, du D_j^2 \{ \sup_{j \in \{1, \ldots, m-1\}} |D_j| \leq m^{-\frac{1}{2} + \alpha} \} \leq cm^{2\alpha - \frac{1}{2}},
\]
again using the exponential martingale inequality, we obtain that for all \( x > 0 \),
\[
P \left( \left| \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) D_j \right| \geq \frac{x}{4} \right) \leq 2 \exp \left( -cm^{\frac{1}{2} - 2\alpha} x^2 \right).
\]
Therefore, choosing \( \alpha > 0 \) such that \( m^{2\alpha} = m^{\frac{3}{4}} x \), we conclude that
\[
P \left( |A_1| \geq \frac{x}{2} \right) \leq 2m \exp \left( -cm^{\frac{3}{4}} x \right).
\]

We next bound \( A_2 \). We write \( A_2 = A_{21} + A_{22} + A_{23} \), where
\[
A_{21} = \sum_{i=1}^{H-1} \int_{t_i}^{t_{i+1}} \theta(u) dB(u) \sum_{j=0}^{i-1} k \left( \frac{i - j}{H} \right) \int_{t_j}^{t_{j+1}} \theta(u) dB(u)
\]
\[
A_{22} = \sum_{i=H}^{m-1} \int_{t_i}^{t_{i+1}} \theta(u) dB(u) \sum_{j=i-H}^{i-1} k \left( \frac{i - j}{H} \right) \int_{t_j}^{t_{j+1}} \theta(u) dB(u)
\]
\[
A_{23} = \sum_{i=m}^{m + H - 1} \int_{t_i}^{t_{i+1}} \theta(u) dB(u) \sum_{j=i-H}^{m-2} k \left( \frac{i - j}{H} \right) \int_{t_j}^{t_{j+1}} \theta(u) dB(u).
\]
We start bounding \( A_{22} \). For a generic \( i \), we have that
\[
\sum_{j=i-H}^{i-1} k^2 \left( \frac{i - j}{H} \right) |\theta(u)|^2 \, du \leq cm^{-\frac{1}{2}}.
\]
Thus, according to the exponential martingale inequality, for any \( \alpha > 0 \), we obtain that

\[
P\left( \sup_{i \in \{H, \ldots, m-1\}} |F_i| > m^{\alpha - \frac{1}{4}} \right) \leq 2m \exp \left(-cm^2\alpha \right),
\]

where \( F_i = \sum_{j=i-H}^{i} \left( \frac{j+1}{2} \right) \int_{t_j}^{t_{j+1}} \theta(u) dB(u) \). Therefore, using again the exponential martingale inequality, we conclude that for any \( x > 0 \) and \( \alpha > 0 \),

\[
P \left( \left| A_{22} \right| \geq \frac{x}{6} \right) \leq P \left( \left| \sum_{i=H}^{m-1} \int_{t_i}^{t_{i+1}} \theta(u) F_i^1 1_{\{ \sup_{i \in \{H, \ldots, m-1\}} |F_i| \leq m^{\alpha - \frac{1}{4}} \}} dB(u) \right| \geq \frac{x}{12} \right) + P \left( \sup_{i \in \{H, \ldots, m-1\}} |F_i| > m^{\alpha - \frac{1}{4}} \right) \leq 2 \exp \left(-cm^{2} - 2\alpha x^2 \right) + 2m \exp \left(-cm^2\alpha \right),
\]

since

\[
\sum_{i=H}^{m-1} \int_{t_i}^{t_{i+1}} |\theta(u)|^2 F_i^2 1_{\{ \sup_{i \in \{H, \ldots, m-1\}} |F_i| \leq m^{\alpha - \frac{1}{4}} \}} du \leq cm^{2\alpha - \frac{1}{2}}.
\]

Thus, choosing \( \alpha \) as in (20), we conclude that \( A_{2,2} \) also satisfies (20). Similarly, so do \( A_{2,1}, A_{2,3}, \) and \( A \).

**Bound in probability of the term \( B \) in (18).** We decompose \( B \) as \( B = B_1 + B_2 \), where

\[
B_1 = \gamma_0(y, u) + \sum_{h=1}^{H} k \left( \frac{h-1}{H} \right) \left( \gamma_h(y, u) + \gamma_{-h}(y, u) \right)
\]
\[
B_2 = \gamma_0(u, y) + \sum_{h=1}^{H} k \left( \frac{h-1}{H} \right) \left( \gamma_h(u, y) + \gamma_{-h}(u, y) \right).
\]

We start bounding \( B_1 \). We write

\[
B_1 = \sum_{j=1}^{t_j} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) G_j,
\]

where

\[
G_j = \sum_{h=1}^{H} k \left( \frac{h-1}{H} \right) \left( u_{j-h} - u_{j-h-1} + u_{j+h} - u_{j+h-1} + u_j - u_{j-1} \right).
\]

Observe that since \( k(0) = 1, k(1) = 0 \), we have

\[
G_j \sim N \left( 0, 2\eta^2 \sum_{h=1}^{H} \left( k \left( \frac{h}{H} \right) - k \left( \frac{h-1}{H} \right) \right)^2 \right).
\]
Moreover, there exists \( \beta_h \in \left[ \frac{h-1}{H}, \frac{h}{H} \right] \) such that
\[
\sum_{h=1}^{H} \left( k \left( \frac{h}{H} \right) - k \left( \frac{h-1}{H} \right) \right)^2 = \frac{1}{H} \sum_{h=1}^{H} \frac{1}{H} k'(\beta_h)^2 \leq cm^{-\frac{1}{2}}.
\]
Therefore, for all \( \alpha > 0 \),
\[
\mathbb{P} \left( \sup_{j \in \{1, \ldots, m-1\}} |G_j| > m^{\alpha - \frac{1}{2}} \right) > 2m \exp \left( -cm^{2\alpha} \right).
\]
Consequently, by the exponential martingale inequality, for any \( x > 0 \),
\[
\mathbb{P} \left( |B_1| \geq x \right) \leq \mathbb{P} \left( \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) \right)_{j \in \{1, \ldots, m-1\}} \left( \sup_{j \in \{1, \ldots, m-1\}} |G_j| \leq m^{\alpha - \frac{1}{2}} \right) \geq x \right) + \mathbb{P} \left( \left| \sum_{j=1}^{m-1} \left( S_j - \mathbb{E} S_j \right) u_{j-1} \right| > m^{\alpha - \frac{1}{2}} \right) \leq 2 \exp \left( -cm^{2\alpha - \frac{1}{2}} \right) + 2m \exp \left( -cm^{2\alpha} \right),
\]
since
\[
\sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u)^2 G_j^2 \left( \sup_{j \in \{1, \ldots, m-1\}} |G_j| \leq m^{\alpha - \frac{1}{2}} \right) du \leq cm^{2\alpha - \frac{1}{2}}.
\]
Thus, choosing \( \alpha \) as in (20), we conclude that \( B_1 \) also satisfies (20).

We next bound \( B_2 \). We write
\[
B_2 = \sum_{j=1}^{m-1} (u_j - u_{j-1}) S_j = -S_1 u_0 + \sum_{j=2}^{m-1} (S_{j-1} - S_j) u_{j-1} + S_{m-1} u_{m-1},
\]
where
\[
S_j = \sum_{h=1}^{H} \left( \frac{h-1}{H} \right) \left( \int_{t_{j-h}}^{t_{j-h-1}} \theta(u) dB(u) + \int_{t_{j+h-1}}^{t_{j+h}} \theta(u) dB(u) + \int_{t_{j-1}}^{t_j} \theta(u) dB(u) \right).
\]
Now, for all \( x > 0 \) and \( \alpha > 0 \), we write
\[
\mathbb{P} \left( |B_2| \geq x \right) \leq \mathbb{P} \left( \left| B_2 \right|_{E \leq m^{2\alpha - \frac{1}{2}}} \geq \frac{x}{2} \right) + \mathbb{P} \left( E > m^{2\alpha - \frac{1}{2}} \right),
\]
where \( E = S_1^2 + \sum_{j=2}^{m-1} (S_{j-1} - S_j)^2 + S_{m-1}^2 \).

Since the random variables \( u_0, u_1, \ldots, u_m \) are iid centered Gaussian with variance \( \eta^2 \),
\[
\mathbb{P} \left( \left| B_2 \right|_{E \leq m^{2\alpha - \frac{1}{2}}} \geq \frac{x}{2} \right) \leq 2 \exp \left( cm^{\frac{1}{2} - 2\alpha} x^2 \right).
\]
In order to bound the second term, we observe that

\[ E = S_1(2S_1 - S_2) + \sum_{j=2}^{m-2} S_j(2S_j - S_{j-1} - S_{j+1}) + S_{m-1}(2S_{m-1} - S_{m-2}). \]

Thus,

\[
P\left( E > 7m^{2\alpha - \frac{1}{2}} \right) \leq P\left( \sup_{j \in \{2, \ldots, m-2\}} |2S_j - S_{j-1} - S_{j+1}| > m^{\alpha - \frac{3}{4}} \right) + P\left( \sup_{j \in \{1, \ldots, m-1\}} |S_j| > m^{\alpha - \frac{5}{4}} \right).
\]

Straightforward computations show that both terms are bounded by \( cm \exp\left(-cm^{2\alpha}\right) \). Therefore, choosing \( \alpha \) as in (20), we conclude that \( B_2 \), and thus \( B \) also satisfies (20).

**Bound in probability of the term \( C \) in (18).** We write

\[ C = -u_0e_1 + \sum_{j=1}^{m-2} (e_j - e_{j+1})u_j + e_{m-1}u_{m-1}, \]

where \( e_j = \sum_{h=2}^{H} k \left( \frac{h-1}{m} \right) (u_{j-h} - u_{j-h-1} + u_{j+h} - u_{j+h-1}) + u_{j+1} - u_{j-2} \).

For all \( x > 0 \) and \( \alpha > 0 \), we write

\[
P\left( |C| \geq x \right) \leq P\left( \left| C_{\{F \leq m^{2\alpha - \frac{1}{2}}\}} \right| \geq \frac{x}{2} \right) + P\left( F > m^{2\alpha - \frac{1}{2}} \right),
\]

where \( F = e_1^2 + \sum_{j=2}^{m-1} (e_j - e_{j-1})^2 + e_{m-1}^2 \).

Again, straightforward computations show that

\[
P\left( F > 3m^{2\alpha - \frac{1}{2}} \right) \leq P\left( \sup_{j \in \{2, \ldots, m-1\}} (e_j - e_{j-1})^2 > m^{2\alpha - \frac{3}{2}} \right) + P\left( e_1^2 > m^{2\alpha - \frac{3}{2}} \right) + P\left( e_{m-1}^2 > m^{2\alpha - \frac{3}{2}} \right) \leq cm \exp\left(-cm^{2\alpha}\right).
\]

Hence, the rest of the proof follows as for \( B_2 \), which shows that \( C \) also satisfies (20).

**Conclusion.** The bounds obtained for \( A, B, \) and \( C \), together with (18) and (19), show that there exist constant \( c_1, c_2, c_3, c_4 > 0 \) such that for large \( m \) and \( x \in [0, c_1] \),

\[
P\left( |\bar{\sigma}_{Kii} - \sigma_{ii}^*| \geq x \right) \leq 2 \exp\left(-c_2mx^2\right) + c_3m \exp\left(-c_4m^{\frac{1}{4}}x\right).
\]

Now, when \( m^{-\frac{3}{4}} < x \leq c_1 \), we have \( x^2m \geq x^{m^{\frac{1}{4}}} \), and thus, the second term wins. On the other
hand, when \(0 \leq x \leq m^{-\frac{3}{4}}\), we have that

\[
m \exp \left( -c_4 x m^{\frac{1}{4}} \right) \geq m \exp \left( -c_4 m^{-\frac{3}{4}} \right) > 1,
\]

and thus the same bound holds. This concludes the proof of the theorem for the diagonal terms.

We are now going to prove the concentration inequality for two generic assets \(i\) and \(j\). By (1), the integrated covariance of the log-returns of assets \(i\) and \(j\) on \([0, 1]\) is given by

\[
\int_0^1 \Theta_i(u)\Theta_j(u)\,du = \frac{1}{4} \left( \int_0^1 (\Theta_i(u) + \Theta_j(u)) (\Theta_i(u) + \Theta_j(u))' \,du \right)
- \frac{1}{4} \left( \int_0^1 (\Theta_i(u) - \Theta_j(u)) (\Theta_i(u) - \Theta_j(u))' \,du \right),
\]

(22)

where \(\Theta_k(u), k \in \{1, \ldots, n\}\) denote the rows of the matrix \(\Theta(u)\). Therefore, in order to estimate the integrated covariance, it suffices to estimate the two terms in (22).

The MRK estimator of \(s_{ij}^* = \int_0^1 (\Theta_i(u) + \Theta_j(u)) (\Theta_i(u) + \Theta_j(u))' \,du\) is given by

\[
s_{Kij} = \gamma_0(x_i^r + x_j^r) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(x_i^r + x_j^r) + \gamma_{-h}(x_i^r + x_j^r)),
\]

where \(\gamma_h(x_i^r + x_j^r) = \gamma_h(x_i^r + x_j^r, x_i^r + x_j^r), \text{ for all } h \in \{-H, \ldots, H\}\).

Observe that if there is no asynchronicity among the observations, then we can derive the concentration inequality for \(s_{Kij}\) as we did for the case of a single asset. Therefore, it suffices to split the estimator as \(s_{Kij} = s_{Kij} + \sum_{q=1}^{10} F_q\), where \(s_{Kij}\) is the analogous of \(s_{Kij}\) but replacing \(x_i^r + x_j^r\) by \(x_i + x_j\), where \(x_{\ell k} := y_{\ell}(\tau_k) + u_{\ell k}^r\), for \(\ell = i, j\) and \(k \in \{1, \ldots, m\}\). The terms \(F_q\), are those that contain the points \(\triangle y_{\ell k} := y_{\ell}(\tau_k) - y_{\ell k}^r\), that is,

\[
F_1 = \gamma_0(\triangle y_i, x_i + x_j) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(\triangle y_i, x_i + x_j) + \gamma_{-h}(\triangle y_i, x_i + x_j)),
\]

\[
F_3 = \gamma_0(x_i + x_j, \triangle y_i) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(x_i + x_j, \triangle y_i) + \gamma_{-h}(x_i + x_j, \triangle y_i)),
\]

\[
F_5 = \gamma_0(\triangle y_i, \triangle y_j) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(\triangle y_i, \triangle y_j) + \gamma_{-h}(\triangle y_i, \triangle y_j)),
\]

\[
F_7 = \gamma_0(\triangle y_i, u_i^r + u_j^r) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(\triangle y_i, u_i^r) + \gamma_{-h}(\triangle y_i, u_i^r + u_j^r)),
\]

\[
F_9 = \gamma_0(u_i^r + u_j^r, \triangle y_i) + \sum_{h=1}^H k \left( \frac{h - 1}{H} \right) (\gamma_h(u_i^r + u_j^r, \triangle y_i) + \gamma_{-h}(u_i^r + u_j^r, \triangle y_i)).
\]
$F_2, F_4, F_6, F_8$ and $F_{10}$ are equal to $F_1, F_3, F_5, F_7$ and $F_9$, respectively, by replacing $\Delta y_i$ by $\Delta y_j$ and viceversa.

The proof of the concentration inequality for a single asset $i$ shows that $|\hat{\sigma}_{K_{ij}} - \sigma_{ij}^*|$ satisfies (20). Straightforward computations show that the same bound is satisfied for all $|F_q|$. Finally, proceeding similarly, we obtain the same estimate for the MRK estimator of the second term in (22), which concludes the desired proof.