High Dimensional Multivariate Realized Volatility Measures

Tim Bollerslev [*], Nour Meddahi[†], Serge Nyawa [‡]

Abstract

We provide a factor based estimator of the covolatility matrix when the number of stocks involved is large and when the recorded data are contaminated by microstructure noises. A statistical factor specification of returns is assumed and evidences of a factor structure in microstructure noises are provided. The estimation approach uses the intuition of the asymptotic Principal Component Analysis of Connor and Korajzcyk (1998). We find that a factor structure in returns and in microstructure noise improves the estimation of the covolatility matrix, the correlation matrix and the inverse of the covolatility matrix, when the number of assets is large. More precisely, under the synchronicity assumption of asset prices, we prove analytically that our estimator outperforms the more popular ones, namely the multirealized kernel of Barndorff-Nielsen et al. (2008), the adjusted Modulated Realized Covariance of Christensen et al. (2010) and the composite kernel of Lunde et al. (2011). Through an intensive simulation exercise, we check that even with asynchronous asset prices, our approach works well. Using a four years TAQ database of the Wharton Research Data Services, we conduct empirical studies in order to compare the risk associated to each estimator in a portfolio allocation exercise, and under gross exposure constraints. The results are in favor of our approach.

Keywords: Covolatility matrix; High-frequency data; Microstructure noise; Robust covolatility measures; High dimensional matrix.

JEL Classification: C13, C32, C58.

[*]Department of Economics, Duke University, NBER and CREATES

[†]Toulouse School of Economics, France

[‡]Toulouse School of Economics, France

1 Introduction

The covariance matrix of asset returns has a huge importance in areas of risk management, portfolio allocation, hedging and asset pricing. More specifically, the solution to the optimal portfolio allocation problem is a function of the inverse of the asset returns covariance matrix, and the risk associated to a given portfolio is related to the covariance matrix of the corresponding asset returns. Hence, having an accurate and well conditioned estimate of the covariance matrix is crucial. In the Financial Econometrics literature, the covariance between asset returns is usually called the *covolatility matrix* and is a latent variable. When high frequency data prices are available for a trading day and generated by a brownian semi-martingale process, the covolatility matrix of asset returns can be obtained through the integral of the spot covariance matrix, and is called *the integrated covolatility matrix*.

The attention to the case of a huge number of assets is recent (see, e.g. Zheng et al. (2011)). This special case is important because in practice, the portfolio allocation or the risk management can involve a large number of assets[1]. The number of assets to take into account can be of the same order or bigger than the sample size (see, e.g. Fan et al. (2008)). However, a direct modeling of the integrated covolatility matrix with a large dimension leads to the estimation of too many parameters. As an example, we face this problem when we use a multidimensional $GARCH$ specification or a diffusion stochastic volatility model. The same problem arises if we estimate each component of the covolatility matrix for a huge number of assets, using the different unidimensional available methods. In this later case, with $p = 100$ assets for example, the number of parameters of the covolatility matrix to estimate will be 5050. But, the number of observations available for each asset is often in the order of hundred. In addition, an estimation point by point doesn't guarantee the crucial property of positive semi-definiteness. As mentioned in the seminal paper of Fan et al. (2008), one solution to these problems consists on imposing a structure to the model.

In the the covolatility matrix estimation literature, there are several ways to impose a structure to the model. The main approaches are the *sparsity*, the *decay* and the *factorial* assumptions. Sparsity condition assumes that the high dimensional covolatility matrix can be for example a block diagonal matrix, a matrix with a relatively small number of

[1]For Statman (1987), a well-diversified portfolio contain at least 40 stocks

nonzero elements in each row or column, or a matrix obtained by randomly permuting rows and columns of above matrices. Under the decay assumption, matrix elements are assumed to decay away from the diagonal. Important terms are within a band. The decay condition is less realistic than the sparsity assumption. In this study, we focus on the factorial assumption. As in the single index model of Sharpe (1994) and Lintner (1965), and as in the arbitrage pricing theory of Ross (1976), we assume that the frictionless asset returns have a factor structure. Our choice of this factorial representation is motivated by many advantages: i) the factorial structure of returns guarantees a semi-definite positive estimator of the covolatility, so we don't need additional regularization of the estimator; ii) the estimator provided in this specification is invertible under weak conditions; iii) the number of parameters to estimate is relatively small; iv) this approach is widely used both theoretically and empirically in economics and finance (see, e.g. Fan et al. (2008)). The idea in our setup will be the introduction of a factor specification in the returns modeling, in order to reduce the number of parameters to estimate, and to guarantee a semi-definite positive and invertible estimator of the covolatility matrix.

There are two main types of factor decomposition of asset returns: the macro-economic factor model (see for example Chen et al. (1986)) and the statistical factor model (see for example Ross (1976)). The difference between the two come from the assumptions on the loading factors and on the common factors. In the macro-economic factor model, factors are observable and obtained from macroeconomic or asset pricing theory while the loading factors are estimated. In the statistical factor model, none the loading factors or the common factor are specified. Both of them are estimated using the returns data.

There are some papers related to the estimation of a high dimensional covolatility matrix through a macro-economic factor model. But in our knowledge, the case of a statistical factor model of returns with prices contaminated by microstructure noises (bid-ask bounds, transaction prices, non-trading periods or price discreteness, trades occurring on different markets or networks, rounding errors...) is not yet studied. In this paper, we want to avoid one important drawback of the macro-economic factor model: difficulties to choose and to measure the right factors. With the statistical factor model, we avoid misspecification and missing factors problems.

To summarize, our estimator of the integrated covariation matrix will be derived under microstructure noises, and using a statistical factor model of returns. Additionally, we will impose a factor structure in the microstructure noise. We will take advantage of both the factor representation and the properties of the pre-averaging estimator of Jacod et al. (2009) in order to construct our estimator under the high dimensional assumption.

The general setup

As it is a common fact in finance, we assume that the data generating process of the *p-dimensional* frictionless log-prices X_t follows a continuous Itô semimartigale process

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{1}$$

where $X_t = \left(X_t^{(1)}, ..., X_t^{(p)}\right)'$ is the true and latent log-price vector; $B_t = \left(B_t^{(1)}, ..., B_t^{(p)}\right)'$ is a *p-dimensional* standard Brownian motion ($B^{(i)}$, $i = 1, ..., p$ are independent standard brownian motion); $\mu_t = \left(\mu_t^{(1)}, ..., \mu_t^{(p)}\right)'$ and $\sigma_t = \left(\sigma_t^{(1)}, ..., \sigma_t^{(p)}\right)'$ are respectively a *p-dimensional* predictable locally bounded drift process and a càdlàg $p \times p$ spot covolatility process.

The object of interest is the integrated volatility matrix of the true price (also called the integrated covariance, integrated covolatility, or integrated covariation matrix)

$$ICV = \int_0^1 \sigma_s \sigma_s' ds \tag{2}$$

It is a measure of the ex-post covariation of asset prices over a trading period of length 1. This quantity is important for portfolio management, asset pricing or forecasting of the next day's integrated covariation, etc.

The rest of the work is structured as follow. In section 2, we present a literature review. Section 3 describes the model and the estimation methodology. Section 4 concerns sampling properties and simulations under the synchronicity assumption. In section 5, we present the asynchronous case. An empirical study is carried out in section 6 and section 7 concludes the paper. All the main proofs are deferred to the technical Appendix.

2 Literature review

In the univariate case, the realized variance is the best estimator of the integrated volatility, when there is not microstructure noise. It is defined by

$$RV_T = \sum_{t_i} (Y_{t_{i+1}} - Y_{t_i})^2 \tag{3}$$

where Y_t is the observed log-price at time t. It has a nice convergence property to the integrated volatility as the number of observation goes to infinity, established by Jacod and Protter (1998). However, under the realistic assumption of microstructure noise, this estimator is not consistent. Some popular solutions of this inconsistency are available. The first is the subsampling and averaging approach of Zhang et al. (2005), and provide the averaging and two scales estimators. The second is the realized kernels of Barndorff-Nielsen et al. (2008). And the third is the pre-averaging approach of Podolskij et al. (2009). The two last approaches will be intensively used in this work.

The kernel estimator is presented below in a general framework. The idea of the pre-averaging approach of Mark Podolskij et al. (2009) is to choose a window of length k_n, a weigthing function g, and to construct from the initial return series a new one by averaging overlapping blocs of $k_n - 1$ consecutive returns. It is with this latter series that the pre-averaging realized variance is construted. The function g has to be continuous with piecewise continuous and Lipschitz derivatives. Jacod et al. (2009) proposed as consistent estimator of the integrated volatility, using the observed return process r, the following pre-averaging estimator

$$PRV(r) = \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{\lfloor 1/\Delta_n \rfloor - k_n + 1} (\overline{Y}_i^n)^2 - \frac{\psi_1 \Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} r_i^2 \tag{4}$$

where n is the number of observed returns; Δ_n is the time interval between two observations; $r_i = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ is the i^{th} return computed from the observed price series Y; $\overline{Y}_i^n = \sum_{j=1}^{k_n-1} g(j/n) r_{i+j}$ is the i^{th} pre-averaging return and θ is a setting parameter to choose optimally such that $k_n\sqrt{\Delta_n} = \theta + o(\Delta_n^{1/4})$. Also $\phi_1(s) = \int_s^1 g'(u)g'(u-s)du$, $\phi_2(s) = \int_s^1 g(u)g(u-s)du$, and $\psi_{\mathrm{i}} = \phi_i(0)$. The most important result of the pre-averaging approach is resumed in the asymptotic behavior established in Jacod et al.(2009).

$$\Delta_n^{-1/4}(PRV(r) - IV) \to N(0;\Gamma) \tag{5}$$

with $\Gamma = \int\limits_0^1 \frac{4}{\psi_2^2}\left(\Phi_{22}\theta\sigma_t^4 + 2\Phi_{12}\frac{\sigma_t^2 V_\epsilon}{\theta} + \Phi_{11}\frac{V_\epsilon^2}{\theta^3}\right)dt$, V_ϵ is the noise variance, IV the true integrated volatility and $\Phi_{ij} = \int\limits_s^1 \phi_i(s)\phi_j(s)ds$.

All the previous estimators work well in the univariate case. When we are interested on more than one asset, there is a need of a multidimensional approach. These estimators can be extended to the multivariate case, but only if the observations of all the different assets are synchronous, it means recorded exactly at the same time (Christensen et al. (2011)), and if the number of assets is small relatively to the sample size. The first assumption is not really realistic, because very often, prices are not recorded at the same time for all the assets. In this case, there are two types of problems to deal with: the first is the microstructure noise and the second the non-synchronicity of observations. When these problems exist, the usual estimators of the covolatility matrix are seriously biased.

To provide a solution to the non-synchronicity problem when estimating the covolatility matrix, Hayashi et al. (2005) propose an estimator of the covariation of two diffusion processes when they are observed only in discrete time. Their estimator is based on overlap intervals and is free of any synchronization process of the original data. However, the estimator of Hayashi et al. (2005) doesn't deal with the microstructure noise.

Thanks to the multivariate realized kernel estimator of Barndorff-Nielsen et al. (2008). These authors construct the first estimator which guarantees simultaneously: consistency, positive semi-definiteness, robust to microstructure error, and handles non-synchronous trading. The non-synchronicity issue is resolved using the refresh time approach. The different steps in the construction of this estimator are the following:

- *Step1:* They deal with the non-synchronicity issue by constructing a new series of synchronous data using the *refresh time* approach. The idea of the refresh time is to wait until all assets are traded at least once at time v_1 (say) and then use the last price traded before or at v_1 of each asset as its price at time v_1 . This yields one synchronized price vector at time v_1. The clock now starts again. Wait until all assets are traded at least once at time v_2 (say) and again use the previous tick price of each asset as its price at time v_2. This yields the second synchronized price vector at time v_2. Repeat the process until all available trading data are synchronized (Fan et al. (2012)).

- *Step2:* There is a need to average m returns (practically m is around 2) at the very beginning and the end of the day to obtain a consistent estimator.

- *Step3:* The formula of the realized kernel is

$$K(Y) = \sum_{h=-n}^{n} k(\frac{h}{H+1})\Gamma_h, \tag{6}$$

$$\Gamma_h = \sum_{j=h+1}^{n} y_j y_{j-h}^{'}, \text{ for } h > 0; \; \Gamma_h = \Gamma_{-h}^{'}, \text{ for } h < 0$$

where n is the number of synchronized returns per asset, Γ_h is the h^{th} realized auto-covariance; $y_j = Y_j - Y_{j-1}$ for $j = 1, 2, ..., n$; with $Y_0 = \frac{1}{m}\sum_{j=1}^{m} Y(\tau_{p,j})$; $Y_n = \frac{1}{m}\sum_{j=1}^{m} Y(\tau_{p,p-m+j})$; $Y_j = Y(\tau_{p,j+m})$ for $j = 1, ..., n-1$; $\{\tau_{p,j}\}$ is the series of refresh time ; and k is a non-stochastic weighting function. The rate of convergence of this estimator is $n^{-1/5}$.

Christensen et al. (2010) propose two estimators for the covolatility matrix of continuous Itô semimartingales, observed with noise on synchronized data. The first is the Modulated Realized Covariance estimator (henceforth MRC) which is a multivariate version of the one dimensional pre-averaging estimator (henceforth PRV). It has an optimal rate of convergence ($n^{-1/4}$), and is unbiased but not necessarily semi-definite positive. it is defined by

$$MRC\left[Y\right]_n = \frac{n}{(n-k_n+2)}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1} \bar{Y}_i^n \left(\bar{Y}_i^n\right)^{'} - \frac{\psi_1^{k_n}}{2n\theta^2\psi_2^{k_n}}\sum_{i=1}^{n}(r_i)(r_i)^{'} \tag{7}$$

where Y is the observed price vector, n is the number of observed returns per asset, $\bar{Y}_i$ the i^{th} averaged return vector, r_i the i^{th} usual return vector defined as in (4), g a weighting function, $\psi_1^{k_n} = k_n \sum_{i=1}^{k_n-1}\left(g(\frac{i}{k_n}) - g(\frac{i-1}{k_n})\right)^2$, $\psi_2^{k_n} = \frac{1}{k_n}\sum_{i=1}^{k_n-1} g^2(\frac{i}{k_n})$, $k_n - 1$ the number of returns in each average, such that $\frac{k_n}{n^{1/2}} = \theta + o(n^{-1/4})$ and θ is a setting parameter. When the assets are not observed at the same time, the non-synchronicity issue is resolved using the refresh time method of Barndorff-Nielsen et al. (2008). The second estimator is a semi-definite positive version of the MRC, the adjusted Modulated Realized Covariance (henceforth MRC^δ). MRC^δ is applied to the aligned series of returns (After a refresh time exercise). Its formula is the following

$$MRC\,[Y]_n^\delta = \frac{n}{(n-k_n+2)}\frac{1}{\psi_2 k_n}\sum_{i=0}^{k_n}\bar{Y}_i^n\left(\bar{Y}_i^n\right)' \tag{8}$$

θ is such that $\frac{k_n}{n^{1/2+\delta}} = \theta + o(n^{-1/4+\delta/2})$. This estimator is consistent, with a sub-optimal rate of convergence of $n^{-1/5}$, and is positive semi-definite.

The above covolatility estimators perform well when the number of assets is small relatively to the sample size. In reality, the number of assets to take into account can be large. In this case, they perform poorly because of the lack of accuracy in estimating high-dimensional matrices (Minjing Tao et Al. (2011)). Two groups of estimators have been proposed in order to resolve the dimensionality issue: those without factors and the approach with factors.

In a setting without the factor representation, Wang et al.(2010) propose a new type of estimators of the integrated covolatility matrix and establish their asymptotic theory under the sparsity or the decay assumption. Zheng et al.(2011) focus on the case where the log-price follows a diffusion process with some particular features: the spot covolatility process is càdlàg and can be written as $\Theta_t = \gamma_t \Lambda$, with Λ a $p \times p$ matrix satisfying $tr(\Lambda\Lambda^T) = p$ and γ_t is a càdlàg and random function. Their approach is to study the estimation of the integrated covariance through the estimation of its spectral distribution. Lunde et al.(2011) construct a composite realized kernel (Henceforth $\hat{\Sigma}_{comp}$) to estimate a high dimensional covolatility matrix. They estimate the integrated covolatilities between the different couples of assets through the bivariate realized kernel. For each couple of assets, the bandwidth H is selected optimally. Then, the resulting matrix is regularized in order to obtain their definite positive estimator $\hat{\Sigma}_{comp}$.

The main part of works based on the estimation of a large covolatility matrix is through the choice of a structure for the covolatility matrix. When the problem concerns the stock returns, the factor representation seems natural (see, e.g. Ledoit et al. (2001)). The idea is to deal with the dimensionality issue and to force the estimator to be well-conditioned (estimation error is not amplified by inverting).

The use of the factor representation to estimate the covolatility is not recent. The main papers in this topic were not focused on the estimation of the integrated volatility, but the identification of global models with a time varying volatility.

In 1989, Nerlov and Diebold proposed to study the temporal volatility patterns in seven major nominal dollar spot exchange rates. Their multivariate approach was based on a latent factor structure to reduce substantially the number of parameters to estimate. The estimation procedure consisted firstly to extract the time series of common factor values, and to test for the model $ARCH$ effect on this factor series. Secondly, the parameters of the model are simultaneously estimated through a multivariate log-likelihood.

In the field of the modelization of the volatility in the high dimensional case, Shephard et al.(2006) proposed a unified Bayesian fitting and inference framework for truly high dimensional multivariate Stochastic Volatility models. Their model is a generalization of a univariate stochastic volatility model, with unobserved factors. They use a Markov Chain Monte Carlo (MCMC) method to develop a practical Bayesian estimation approach (Chib and Greenberg (1996) and Chib (2001)).

Fan et al. (2008) examine how the dimensionality impact the estimation of the covariance matrices. They use a multi-factor model for the vector of excessive returns of p assets to resolve the problem due to the dimensionality and to estimate the covariance matrix[2]. Their factors are assumed to be observable. They study the properties of this factorial covariance estimator and make a comparison analysis with the usual sample estimator. Their factorial estimator seems to perform well in estimations which involve the inverse of the covariance matrix. But concerning the estimation of the covariance matrix itself, the gain due to the use of the factor approach is not substantial.

The papers more related to our topic are the one of Tao et al. (2011) and Bannouh et al. (2012). The two papers propose two different approaches to combine low-frequency and high-frequency data in order to estimate the integrated covolatility matrix in the high dimension framework.

In the first paper, the idea is to estimate firstly the daily realized covolatility matrices (for L days) based on high frequency data[3]. Secondly, they fit a matrix factor model to the L estimated daily covolatility matrices. Then, they fit a vector autoregressive model for the estimated volatility factor. The last step consists to build the dynamic structure of the volatility matrix, using the previous VAR modelization of the factor volatility matrix.

Bannouh et al. (2012) introduce a Mixed-Frequency Factor Model (MFFM) to esti-

[2]motivated by the Arbitrage Pricing Theory in finance, e.g: Factor model of Fama

[3]This is done using the thresholding average realized volatility matrix of Wang and Zou (2010)

mate the vast covolatility matrix of asset returns. They consider as factors highly liquid assets such as exchange traded funds (ETFs) and use these very high-frequency data to estimate the covolatility matrices of the observed factor components. The idiosyncratic risk covolatility matrix is obtained after the regression of the observed returns on the observed factors, at a lower frequency: it is the estimated residuals variances on the diagonal and zeroes elsewhere. Loading factors are estimated using daily data of returns and factors. The main difference with our work is that, here, factors are observed.

3 Model and estimation methodology

3.1 The benchmark model

We assume that the continuous Itô semimartingale process X_t described in (1) follows a factor model of the form

$$dX_t = bdF_t + dE_t \tag{9}$$

where $b = (b_{ik})_{1\leq i\leq p, 1\leq k\leq K}$ is the $p \times K$ matrix of loadings; $F_t = (F_{1t}, ..., F_{Kt})'$ the vector of factor components at time t, and $E_t = (E_{1t}, ..., E_{pt})'$ the idiosyncratic error vector at time t; F and E are uncorrelated; the components of F are uncorrelated each others, same for the components of E. More over,

$$dF_{kt} = \sigma^2_{fkt} dB^F_{kt} \text{ and } dE_{it} = \sigma^2_{\epsilon it} dB^E_{it}$$

We are going to focus on the discrete time version of this model, since the recorded data are discrete and the continuous time model can be obtained by taking the recorded time interval to zero. Initially, we assume that all assets are recorded at equidistant time intervals Δ, in a synchonous way, such that each asset log-price is recorded at $n + 1$ discrete times $0, \Delta, ..., n\Delta$. Since random and asynchounous transaction times are important characteristics of high frequency transaction data, our estimator will be transformed to take into account this situation.

We assume that there are p assets. The asset i has a frictionless log-price at time t denoted p^*_{it}. The frictionless return of asset i over a period $[t - \Delta, t]$ is given by

$$r^*_{it} = p^*_{it} - p^*_{it-\Delta} \tag{10}$$

We assume that there are K factors $f_1, ..., f_K$ such that, $\forall i = 1, ..., p$ the i^{th} frictionless return satisfies

$$r_{it}^* = \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \tag{11}$$

Using a matrix notation

$$r_t^* = b f_t + \varepsilon_t \tag{12}$$

with $r^* = (r_1^*, ..., r_p^*)'$; $b = (b_{ik})_{1 \le i \le p, 1 \le k \le K}$ the $p \times K$ matrix of loadings; $f_t = (f_{1t}, ..., f_{Kt})'$ the vector of factor components at time t, and $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{pt})'$ the idiosyncratic error vector at time t.

As presented previously, the following assumptions are satisfied:

Assumption 1

- $\forall i \in \{1, ..., p\}$, $\forall k \in \{1, ..., K\}$,$Cov\left(f_{kt}, \varepsilon_{it} | I_{t-\Delta}\right) = 0$;
- $Cov\left(f_{kt}, f_{k't} | I_{t-\Delta}\right) = 0$, $\forall k \neq k'$;
- $\forall i \neq j$, $Cov(\varepsilon_{it}, \varepsilon_{jt} | I_{t-\Delta}) = 0$;
- $E\left(\varepsilon_{it} | I_{t-\Delta}\right) = 0$.

I_t is the set of information available at time t.

The use of a factor model is common in the theoretical and empirical literature since the work of Ross S.A (1976) and Chamberlain and Rothschild (1983). The factor component f_t can be seen as general influences which tend to affect all returns. We assume that loading coefficients b_{ik}, $1 \le i \le p$ and $1 \le k \le K$ are not time varying and do not depend on Δ. The observed price p_{it} of asset i is assumed to contain the microstructure noise u_{it}, such that:

$$p_{it} = p_{it}^* + u_{it} \tag{13}$$

We suppose that the microstructure noise has also a factor representation, that is

$$u_{it} = c_i g_t + \eta_{it} \tag{14}$$

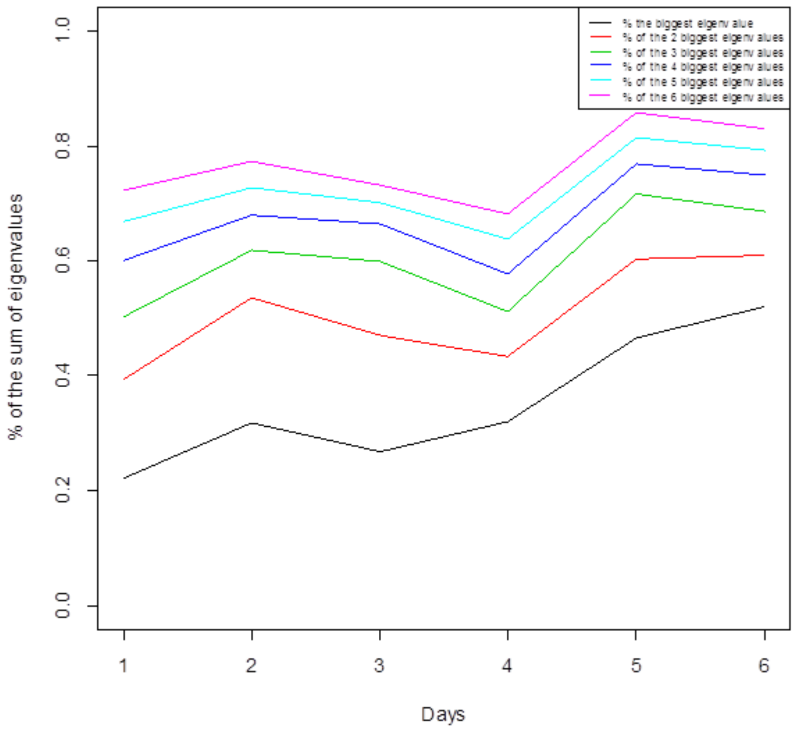
Assumption 2

- c_i is a $1 \times K'$ vector of noise loadings;

- g_t is $K' \times 1$ and captures the cross-sectional dependence in the measurement errors, it is independent across time and independent with f_t and ε_{it};
- η_{it} is supposed to be independent across assets and across time, independent with f_t, ε_{it} and g_t;
- $Var(\eta_{it}) = \sigma^2_{\eta i}$ is constant;
- $Var(g_t) = \sigma^2_g$ is constant.

In order to provide an empirical evidence of the factor structure of microstructure noises, we run the following experiment: we consider a data set of 384 stocks traded on $S\&P500$ and we focus on six trading days, say 11/01/2006, 12/01/2006, 13/06/2008, 16/06/2008, 04/11/2011 and 07/11/2011. For each of these trading days, we compute the realized covariance matrix and we divide it by $2n$, where n is the number of intraday transaction times after synchronization. By doing so, we get an estimator of the covolatility of the microstructure noise. The next step consists on a spectral decomposition of the obtained matrix. The following figure plots the ratio of the sum of the largest eigenvalues (the biggest eigenvalue, the first two biggest eigenvalues, the first three biggest eigenvalues, until the first six biggest eigenvalues) to the total sum of eigenvalues: these ratios can been interpreted as the part of the total variability explained by the considered factors (the first factor, the first two factors, until the first six factors).

Figure 1. Ratio of first biggest eigenvalues relative to the total sum of eigenvalues



From the graphic above, we can see that the four biggest eigenvalues of the microstructure noise covolatility matrix explain more than 60% of the total variability. This is an evidence of the factorial structure of microstructure noise.

The object of interest here remain the integrated covolatility matrix of the frictionless price in (9), defined by

$$\Sigma = bDiag\left[\int_0^1 \sigma_{f1u}^2 du, ..., \int_0^1 \sigma_{fKu}^2 du\right] b' + Diag\left[\int_0^1 \sigma_{\varepsilon 1u}^2 du, ..., \int_0^1 \sigma_{\varepsilon pu}^2 du\right] \tag{15}$$

We want to estimate the previous matrix in the high dimensional case[4]. An intuitive estimator is obtained by plug in the estimators of b, $\int_0^1 \sigma_{fku}^2 du$, and $\int_0^1 \sigma_{\varepsilon iu}^2 du$, $\forall i = 1, ..., p$. Thus, the parameters of interest are: $\int_0^1 \sigma_{fku}^2 du$, $\int_0^1 \sigma_{\varepsilon iu}^2 du$, and b_{ik}, $i = 1, ..., p$ and $k = 1, ..., K$.

[4] p is sufficiently large and can be of the order of n

3.2 Estimation methodology

The following notations will be intensively used in the theoretical analysis:

- $z_{\cdot t} = \frac{1}{p}\sum_{i=1}^{p} z_{it}$;
- $\overline{a}_p = \frac{1}{p}\sum_{i=1}^{p} a_i$;
- $MRC(p_i, p_j)$: the modulated realized covariance estimator of the covariation between the latent prices p_i^* and p_j^*, applied to the observed prices p_i and p_j as defined in (7);
- $PRV(r)$: the pre-averaging estimator of the integrated volatility of the true price p^*, applied to the observed return r as defined in (4);
- $\left[r_i^*, r_j^*\right]$: the true integrated covariation between the prices p_i^* and p_j^* with corresponding returns r_i^* and r_j^*;
- $[X]$: the integrated volatility of the process X;
- $\hat{X}$ and X^{ϵ}: respectively the estimate and the estimation error of the parameter X.

We are going to take advantage of the factor analysis literature and the pre-averaging approach described in (4), (7) and (8) to estimate all our parameters of interest. Firstly, we are going to assume that the observation frequency Δ is constant over time and over assets, and closed to zero.Through the factor analysis methodology, applied to the latent series r^*, we are going to estimate rotations of factors $\{f_{j\Delta}\}_{(j=1,\dots,\lfloor\frac{1}{\Delta}\rfloor)}$: following the intuition of the asymptotic Principal Component Analysis of Connor and Korajzcyk (1998), they will be obtained by minimizing a scaled sum of squared errors of the idiosyncratic component subject to an identification assumption on loading factors. Secondly, Δ is taken to 0 in order to derive the solution in continuous time. Since the resulting estimator is not feasible (r^* is unobserved), we are going to propose a feasible estimator of rotations of factors f_t based on observed return data. The last step will consist on using our estimate of the rotated factors to derive all the parameters of interest, namely: rotated loadings, integrated volatility of the rotated factors, and integrated volatility of the idiosyncratic term. The pre-averaging approach will be used intensively in this last part.

Remark: Let H be a $K \times K$ orthogonal matrix ($H'H = I_K$). Then

$$
\begin{aligned}
r_t^* &= bf_t + \epsilon_t \\
&= bI_K f_t + \epsilon_t \\
&= bH'Hf_t + \epsilon_t \\
&= \tilde{b}\tilde{f}_t + \epsilon_t, \quad with \ \ \tilde{b} = bH' \ and \ \tilde{f}_t = Hf_t
\end{aligned}
$$

From the last equality, it follows that Σ can also be derived using rotations of factors and loadings

$$
\Sigma = \tilde{b} Diag\left[\int_0^1 \sigma^2_{\tilde{f}1u}du, ..., \int_0^1 \sigma^2_{\tilde{f}Ku}du\right]\tilde{b}' + Diag\left[\int_0^1 \sigma^2_{\varepsilon 1u}du, ..., \int_0^1 \sigma^2_{\varepsilon Nu}du\right] \tag{16}
$$

An estimator of Σ is obtained by plug in estimators of b, $\int_0^1 \sigma^2_{fku}du$, and $\int_0^1 \sigma^2_{\varepsilon iu}du$ (or $\tilde{b}$, $\int_0^1 \sigma^2_{\tilde{f}ku}du$, and $\int_0^1 \sigma^2_{\varepsilon iu}du$) for $i = 1, ..., p$ and $k = 1, ..., K$.

The methodology that we are going to present consists on four main steps:

- Estimation of rotated factors $\tilde{f}$;
- Estimation of the integrated volatility of $\tilde{f}$;
- Estimation of the corresponding loadings $\tilde{b}$;
- Estimation of the integrated volatility of the idiosyncratic component.

3.3 Estimation of a rotation $\tilde{f}$ of the factor component

In this section, we estimate a rotation of the factor component using the intuition of the asymptotic Principal Component Analysis of Connor and Korajzcyk (1998) (Henceforth PCA). We consider the following least squared problem where $f_{j\Delta}$ is chosen in order to minimize the scaled sum of squared values of the idiosyncratic component:

$$
\begin{cases}
\underset{f_{j\Delta},b}{Min} \ \ \frac{1}{p} \sum\limits_{j=1}^{\lfloor 1/\Delta \rfloor} (r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta}) \\
s.t \ \ \frac{1}{p} b'b = I_K
\end{cases}
$$

Let's resolve this optimization problem for $f_{j\Delta}$. After developing the constraints we obtain:

$$
\begin{cases}
\underset{f_{j\Delta},b}{Min} \ \ \frac{1}{p} \sum\limits_{j=1}^{\lfloor 1/\Delta \rfloor} (r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta}) \\
s.t \ \ \forall k = 1, ..., K, \frac{1}{p} \underline{b}'_k \underline{b}_k = 1 \\
\forall k = 1, ..., K, \forall l = k+1, ..., K, \underline{b}'_k \underline{b}_l = 0
\end{cases}
$$

where $\underline{b}_k$ corresponds to the column k of b. The Lagrangian of this problem is defined by

$$L = \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}(r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta}) - \sum_{k=1}^{K}\lambda_k(\underline{b}'_k\underline{b}_k - p) - \sum_{k=1}^{K}\sum_{l=k+1}^{K}\mu_{kl}\underline{b}'_k\underline{b}_l$$

By deriving this Lagrangian with respect to $f_{k\Delta}$, we obtain

$$\begin{aligned}
\frac{\partial L}{\partial f_{k\Delta}} &= \frac{\partial}{\partial f_{k\Delta}}\left[\frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}(r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta})\right] \\
&= \frac{\partial}{\partial f_{k\Delta}}\left[\frac{1}{p}(r^*_{k\Delta} - bf_{k\Delta})'(r^*_{k\Delta} - bf_{k\Delta})\right] \\
&= \frac{\partial}{\partial f_{k\Delta}}\left[\frac{1}{p}(r^{*'}_{k\Delta}r^*_{k\Delta} - r^{*'}_{k\Delta}bf_{k\Delta} - f'_{k\Delta}b'r^*_{k\Delta} + f'_{k\Delta}b'bf_{k\Delta})\right] \\
&= (-b'r^*_{k\Delta} - b'r^*_{k\Delta} + b'bf_{k\Delta} + b'bf_{k\Delta}) \\
&= (-2b'r^*_{k\Delta} + 2b'bf_{k\Delta})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L}{\partial f_{k\Delta}} = 0 &\iff (-2b'r^*_{k\Delta} + 2b'bf_{k\Delta}) = 0 \\
&\iff b'bf_{k\Delta} = b'r^*_{k\Delta} \\
&\iff f_{k\Delta} = (b'b)^{-1}b'r^*_{k\Delta} \\
&\iff f_{k\Delta} = (pI_K)^{-1}b'r^*_{k\Delta} \\
&\iff f_{k\Delta} = \frac{1}{p}b'r^*_{k\Delta}
\end{aligned}$$

Hence,

$$f_{k\Delta} = \frac{1}{p}b'r^*_{k\Delta}, \quad \forall k = 1, ..., \lfloor 1/\Delta \rfloor \tag{17}$$

We are going now to concentrate the objective function by replacing $f_{j\Delta}$ by its formula given by (17).

$$\begin{aligned}
\frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}(r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta}) &= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}(r^*_{j\Delta} - b.\frac{1}{p}b'r^*_{j\Delta})'(r^*_{j\Delta} - b.\frac{1}{p}b'r^*_{j\Delta}) \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}(I_p - \frac{1}{p}bb')'(I_p - \frac{1}{p}bb')r^*_{j\Delta} \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}bb'r^*_{j\Delta} \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}\sum_{k=1}^{K} r^{*'}_{j\Delta}\underline{b}_k\underline{b}'_k r^*_{j\Delta} \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{k=1}^{K}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}\underline{b}_k\underline{b}'_k r^*_{j\Delta} \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{k=1}^{K}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}\left(r^{*'}_{j\Delta}\underline{b}_k\right)\left(\underline{b}'_k r^*_{j\Delta}\right) \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{k=1}^{K}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}\left(\underline{b}'_k r^*_{j\Delta}\right)\left(r'^*_{j\Delta}\underline{b}_k\right) \\
&= \frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^{*'}_{j\Delta}r^*_{j\Delta} - \frac{1}{p}\sum_{k=1}^{K}\underline{b}'_k\left(\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r^*_{j\Delta}r'^*_{j\Delta}\right)\underline{b}_k
\end{aligned}$$

From the last equality, we deduce that the optimal $b = (\underline{b}_1, ..., \underline{b}_K)$ is the solution of the following problem

$$\begin{cases} \underset{\underline{b}_1,...,\underline{b}_K}{Max} \quad \frac{1}{p}\sum_{k=1}^{K} \underline{b}_k' \left(\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r_{j\Delta}^* r_{j\Delta}'^* \right) \underline{b}_k \\ s.t \ \ \forall k = 1, ..., K, \frac{1}{p}\underline{b}_k'\underline{b}_k = 1 \\ \forall k = 1, ..., K, \forall l = k+1, ..., K, \underline{b}_k'\underline{b}_l = 0 \end{cases}$$

The problem above is equivalent to resolve K optimization problems defining by: $\forall k \in \{1, ..., K\}$:

$$\begin{cases} \underset{\underline{b}_k}{Max} \quad \frac{1}{p}\underline{b}_k' \left(\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r_{j\Delta}^* r_{j\Delta}'^* \right) \underline{b}_k \\ s.t \ \ \frac{1}{p}\underline{b}_k'\underline{b}_k = 1 \\ \forall l \neq k, \underline{b}_k'\underline{b}_l = 0 \end{cases} \tag{18}$$

The Lagrangian of the above problem has the following form

$$L \ = \ \frac{1}{p}\underline{b}_k' \left(\sum_{j=1}^{\lfloor 1/\Delta \rfloor} r_{j\Delta}^* r_{j\Delta}'^* \right) \underline{b}_k - \lambda_k \left(\frac{1}{p}\underline{b}_k'\underline{b}_k - 1 \right) - \sum_{l \neq k}^{K} \mu_{kl}\underline{b}_k'\underline{b}_l$$

By resolving for $\underline{b}_k$

$$\frac{\partial L}{\partial \underline{b}_k} \ = \ \frac{2}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r_{j\Delta}^* r_{j\Delta}^{*'} \right] \underline{b}_k - \frac{2\lambda_k}{p}\underline{b}_k - \sum_{l \neq k} \mu_{kl}\underline{b}_l$$

$$\frac{\partial L}{\partial b} = 0 \Longleftrightarrow \frac{2}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r_{j\Delta}^* r_{j\Delta}^{*'} \right] \underline{b}_k - \frac{2\lambda_k}{p}\underline{b}_k - \sum_{l \neq k} \mu_{kl}\underline{b}_l = 0$$

By a left multiplication by $\underline{b}_m'$ ($\forall m \neq k$)

$$\begin{aligned} & \frac{2}{p}\underline{b}_m' \sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r_{j\Delta}^* r_{j\Delta}^{*'} \right] \underline{b}_k - \frac{2\lambda_k}{p}\underline{b}_m'\underline{b}_k - \sum_{l \neq k} \mu_{kl}\underline{b}_m'\underline{b}_l = 0 \\ \Leftrightarrow \ & \frac{2}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \underline{b}_m' \left[r_{j\Delta}^* r_{j\Delta}^{*'} \right] \underline{b}_k - \frac{2\lambda_k}{p}\underline{b}_m'\underline{b}_k - \mu_{km}\underline{b}_m'\underline{b}_m = 0 \\ \Leftrightarrow \ & \mu_{km} = 0 \end{aligned}$$

The third equation comes from the uncorrelation assumption of factors and the identification constraint on loadings. Hence, $\forall m \neq k$, $\mu_{km} = 0$. We deduce that

$$\frac{2}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r_{j\Delta}^* r_{j\Delta}^{*'} \right] \underline{b}_k - \frac{2\lambda_k}{p}\underline{b}_k = 0$$

This is equivalent to

$$\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r^*_{j\Delta} r^{*'}_{j\Delta} \right] \underline{b}_k - \lambda_k \underline{b}_k = 0$$

It follows that $\underline{b}_k$ is an eigenvector associated to the matrix $\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r^*_{j\Delta} r^{*'}_{j\Delta} \right]$.

Let's summarize what we have established:

- the factor component is estimated by

$$\hat{f}_{k\Delta} = \tfrac{1}{p} W' r^*_{k\Delta}, \quad \forall k = 1, ..., \lfloor 1/\Delta \rfloor$$

- W is the matrix of ordered eigenvectors of $\sum_{j=1}^{\lfloor 1/\Delta \rfloor} \left[r^*_{j\Delta} r^{*'}_{j\Delta} \right]$.

By taking $\Delta \to 0$ we deduce that

$$\hat{f}_t = \frac{1}{p} W' r^*_t, \quad \forall t > 0 \tag{19}$$

with columns of W corresponding to ordered eigenvectors of Σ.

Remark: From the optimization problem (17) (b is replace by an unknown W to avoid confusion), if we take $\Delta \to 0$, then the k^{th} column $\underline{W}_k$ of the weighting matrix W is determined in order to maximize the integrated volatility of the factor f_k as described by the following problem

$$\begin{cases} \underset{\underline{W}_k}{Max} \quad \frac{1}{p} \underline{W}'_k \Sigma \underline{W}_k \\ s.t \ \ \frac{1}{p} \underline{W}'_k \underline{W}_k = 1 \\ \forall l \neq k, \underline{W}'_k \underline{W}_l = 0 \end{cases} \tag{20}$$

We want to use this approach of determining W in order to prove the consistency in the estimation of the factor component.

The estimator defined by (19) is not feasible because r^*_t and Σ are latent. In order to obtain a feasible estimator, we need consistent estimators of ordered eigenvectors W of Σ. Let's consider $\hat{W}$ the matrix of K ordered eigenvectors of an estimator $\hat{\Sigma}$ of Σ, robust to microstructure noise, with consistent eigenvectors. The *MRker* is a good candidate for such estimator. In order to confirm this fact, we run the following experiment: we simulate using the same framework described below, daily paths of efficient prices of p assets ($p \in \{50, 100, 300\}$) corresponding to three levels of microstructure noise (the low level, with the noise to signal ratio=0.001, the medium level with the noise to signal ratio=0.01 and the high level with the noise to signal ratio=0.1). Prices are supposed to be generated by a two factors model. For each path we compute the corresponding efficient covolatility matrix, and the $MRker$ and derive their spectral decompositions. The following figure illustrates the result

Figure 2. Eigenvectors estimation using the multirealized kernel $MRker$: the low noise level

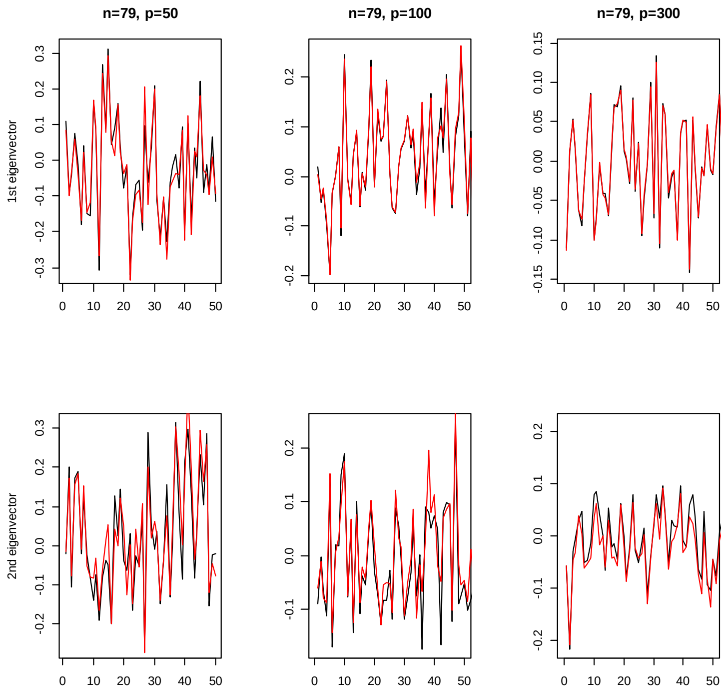


Figure 3. Eigenvectors estimation using the multirealized kernel $MRker$: the medium noise level

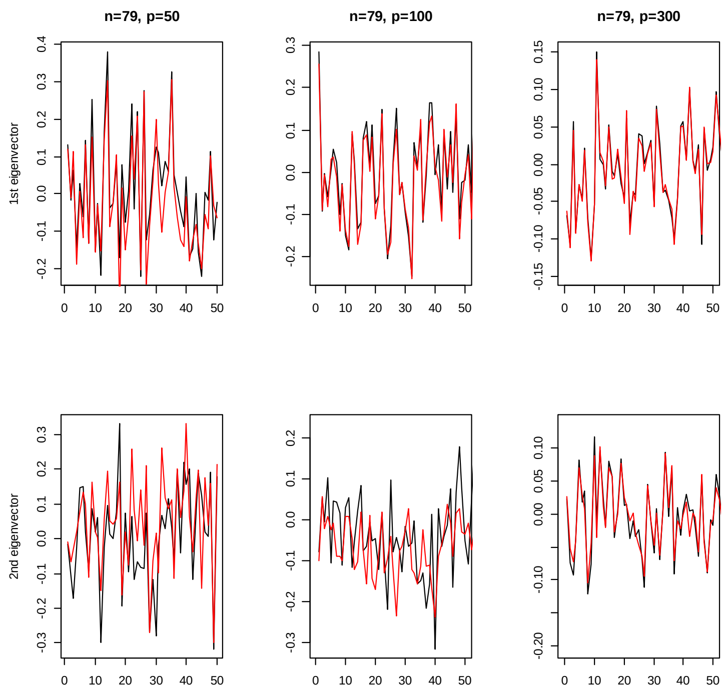
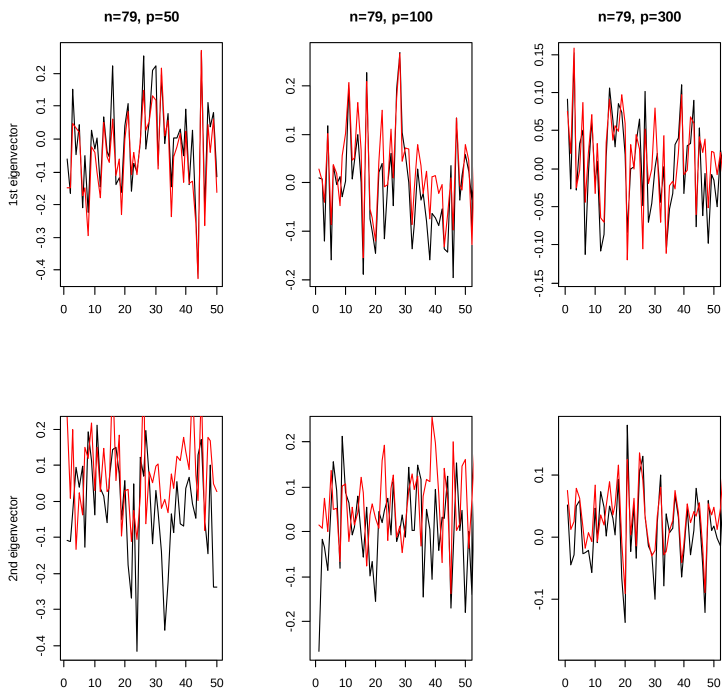


Figure 4. Eigenvectors estimation using the multirealized kernel $MRker$: the high noise level



As we can observe, for a two factor model specification, the two first eigenvectors of the latent covolatility matrix are well estimated by eigenvectors of the $MRker$, independently on the noise level. This observation is in line with the "Linear Shrinkage" estimator of the covariance matrix of LEDOIT et Al.(2002): in order to improve the estimator of the covariance matrix in a large dimensional framework, the "Linear Shrinkage" estimator is obtained from the spectral decomposition of the sample covariance matrix by keeping eigenvectors, while eigenvalues are optimally transformed.

We propose as feasible estimator

$$\hat{f}_t = \frac{1}{p}\hat{W}' r_t \tag{21}$$

where r_t is the series of the $p \times 1$ vector of observed returns, $\hat{W} = \left(\underline{\hat{W}}_1, ..., \underline{\hat{W}}_K\right)$ is a consistent estimator of the $p \times K$ matrix W of ordered eigenvectors of Σ. In practice, we advocate to take as $\hat{W}$ eigenvectors of the $MRker$.

We need now to check that $\hat{f}$ is consistent for the estimation of a rotation $\tilde{f}$ of f plus a microstruture noise component. We firstly express $\hat{f}$ as a function of the true factor f, the idiosyncratic component ϵ_t, and the microstructure noise component u_t

$$\begin{aligned}
\hat{f}_t &= \tfrac{1}{p}\hat{W}' r_t \\
&= \tfrac{1}{p}\hat{W}' \left[r_t^* + (u_t - u_{t-\Delta})\right] \\
&= \tfrac{1}{p}\hat{W}' \left[b f_t + \epsilon_t + (u_t - u_{t-\Delta})\right] \\
&= \tfrac{1}{p}\hat{W}' b f_t + \tfrac{1}{p}\hat{W}'\epsilon_t + \tfrac{1}{p}\hat{W}'(u_t - u_{t-\Delta}) \\
&= \tfrac{1}{p}\hat{W}' b f_t + \tfrac{1}{p}\hat{W}'\epsilon_t + \tfrac{1}{p}\hat{W}' c(g_t - g_{t-\Delta}) + \tfrac{1}{p}\hat{W}'(\eta_t - \eta_{t-\Delta})
\end{aligned}$$

The consistency result in the estimation of a rotation $\tilde{f}$ of f contaminated by a microstructure noise component is given in the following theorem inspired by the paper of Stock and Watson (2002).

Theorem 3.1 *There exists an orthogonal matrix S such that $S\hat{f}$ consistently estimates f. More specifically, as $n = \frac{1}{\Delta}, p \to \infty$:*

- $\frac{1}{p}S\hat{W}' b f_t \xrightarrow{p} f_t$;
- $\frac{1}{p}S\hat{W}'\epsilon_t \xrightarrow{p} 0$;
- $\frac{1}{p}S\hat{W}' c(g_t - g_{t-\Delta})$ *contaminates the rotation of the true factor asymptotically;*
- $\frac{1}{p}S\hat{W}'(\eta_t - \eta_{t-\Delta}) \xrightarrow{p} 0$.

3.4 Estimation of $\int_0^1 \sigma^2_{\tilde{f}ku} du$, $k = 1, ..., K$

We consider the following decomposition of $\hat{f}_t$

$$\hat{f}_{kt} = \frac{1}{p} W_k' r_t^* + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) + \frac{1}{p} W_k^{\epsilon'} r_t^* + \frac{1}{p} W_k^{\epsilon'}(u_t - u_{t-\Delta})$$

We assume that p and n are sufficiently large such that $\frac{1}{p} W_k^{\epsilon'} r_t^*$ and $\frac{1}{p} W_k^{\epsilon'}(u_t - u_{t-\Delta})$ can be neglected. Then

$$\begin{aligned}
\hat{f}_{kt} &= \frac{1}{p} W_k' r_t^* + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) + \frac{1}{p} W_k^{\epsilon'} r_t^* + \frac{1}{p} W_k^{\epsilon'}(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p} W_k' r_t^* + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p} W_k' b f_t + \frac{1}{p} W_k' \epsilon_t + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p} W_k' b f_t + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) \\
&\approx \tilde{f}_{kt} + \frac{1}{p} W_k'(u_t - u_{t-\Delta}) \\
&\approx \tilde{f}_{kt} + \frac{1}{p} W_k' c(g_t - g_{t-\Delta}) + O_p(p^{-1/2} n^{-1/2})
\end{aligned}$$

The last three approximations are derived since $\frac{1}{p} W_k' \epsilon_t = O_p(n^{-1/2} p^{-1/2})$ and $\frac{1}{p} W_k' \eta_t = O_p(n^{-1/2} p^{-1/2})$ and p, n are assume to be large.
We observe $\hat{f}$ looks like a rotation of the latent factor f contaminated by microstructure noises. Hence, by the literature on the estimation of integrated volatility using data contaminated by microstructure noises, $\int_0^1 \sigma^2_{\tilde{f}ku} du$ can be estimated by

$$\widehat{\int_0^1 \sigma^2_{\tilde{f}ku} du} = PRV(\hat{f}_k) \tag{22}$$

3.5 Estimation of the loading factors $\tilde{b}_{ik}$, $i = 1, ..., p$, $k = 1, ..., K$

Since factors are independent pairwise and independent with the idiosyncratic component, it can be established that the integrated covolatility between r_i^* and $\tilde{f}_k$ is equal to $\tilde{b}_{ik}.IV(\tilde{f}_k)$. Thus, we derive that $\tilde{b}_{ik} = \frac{ICV(r_i^*, \tilde{f}_k)}{IV(\tilde{f}_k)}$. Then, an estimator of $\tilde{b}_{ik}$ is deduced as

$$\hat{b}_{ik} = \frac{MRC(r_i, \hat{f}_k)}{PRV(\hat{f}_k)} \tag{23}$$

3.6 Estimation of $\int_0^1 \sigma^2_{\varepsilon iu} du$, $i = 1, ..., p$

We define the variable $\hat{\epsilon}_{it}$ by

$$\hat{\epsilon}_{it} = r_{it} - \sum_{k=1}^{K} \hat{b}_{ik} \cdot \hat{f}_{kt}$$

It is easy to prove that

$$\hat{\epsilon}_{it} = \epsilon_{it} + (u_t - u_{t-\Delta}) - \sum_{k=1}^{K} \tilde{b}_{ik} \tilde{f}_{kt}^{\epsilon} - \sum_{k=1}^{K} \tilde{b}_{ik}^{\epsilon} \tilde{f}_{kt} - \sum_{k=1}^{K} \tilde{b}_{ik}^{\epsilon} \tilde{f}_{kt}^{\epsilon}$$

Since $\hat{f}_{kt}$ is a consistent estimator of $\tilde{f}_{kt}$, let's assume that n and p are sufficiently large such that $\tilde{f}_{kt}^{\epsilon} \equiv \hat{f}_{kt} - \tilde{f}_{kt}$ can be neglected. Then

$$\begin{aligned} \hat{\epsilon}_{it} &\approx \epsilon_{it} + (u_t - u_{t-\Delta}) - \sum_{k=1}^{K} \tilde{b}_{ik}^{\epsilon} \tilde{f}_{kt} \\ &\approx \epsilon_{it} + (u_t - u_{t-\Delta}) + O_p(n^{-3/4}) \\ &\approx \epsilon_{it} + (u_t - u_{t-\Delta}) \end{aligned}$$

It follows that $\hat{\epsilon}_{it}$ is the idiosyncratic component ϵ_{it} contaminated by microstruture noises. Thus, following the literature, we define an estimator of $\int_0^1 \sigma_{\varepsilon iu}^2 du$ for $i = 1, ..., p$ by

$$\widehat{\int_0^1 \sigma_{\varepsilon iu}^2 du} = PRV(\hat{\epsilon}_i) \tag{24}$$

3.6.1 Estimator of the covolatility matrix

We want to recover an estimator of the covariation matrix of $\left(p_1^*, ..., p_p^*\right)$ using the previous estimators. Our estimator $\widehat{\Sigma}$ is obtained by plug in the estimators of $\tilde{b}$, $\int_0^1 \sigma_{fku}^2 du$ for $k = 1, ..., K$ and $\int_0^1 \sigma_{\varepsilon iu}^2 du$, for $i = 1, ..., p$ in (16). We obtain

$$\begin{aligned} \widehat{\Sigma} &= \begin{pmatrix} \hat{b}_{11} & \cdots & \hat{b}_{1K} \\ \vdots & & \vdots \\ \hat{b}_{p1} & \cdots & \hat{b}_{p1} \end{pmatrix} \begin{pmatrix} \widehat{\int_0^1 \sigma_{f1u}^2 du} & & \\ & \ddots & \\ & & \widehat{\int_0^1 \sigma_{fKu}^2 du} \end{pmatrix} \begin{pmatrix} \hat{b}_{11} & \cdots & \hat{b}_{p1} \\ \vdots & & \vdots \\ \hat{b}_{1K} & \cdots & \hat{b}_{pK} \end{pmatrix} \\ &\quad + \begin{pmatrix} \widehat{\int_0^1 \sigma_{\varepsilon 1u}^2 du} & & \\ & \ddots & \\ & & \widehat{\int_0^1 \sigma_{\varepsilon pu}^2 du} \end{pmatrix} \\ &= \begin{pmatrix} \frac{MRC(r_1, \hat{f}_1)}{PRV(\hat{f}_1)} & \cdots & \frac{MRC(r_1, \hat{f}_K)}{PRV(\hat{f}_K)} \\ \vdots & & \vdots \\ \frac{MRC(r_p, \hat{f}_1)}{PRV(\hat{f}_1)} & \cdots & \frac{MRC(r_p, \hat{f}_K)}{PRV(\hat{f}_K)} \end{pmatrix} \begin{pmatrix} PRV(\hat{f}_1) & & \\ & \ddots & \\ & & PRV(\hat{f}_K) \end{pmatrix} \\ &\quad \begin{pmatrix} \frac{MRC(r_1, \hat{f}_1)}{PRV(\hat{f}_1)} & \cdots & \frac{MRC(r_p, \hat{f}_1)}{PRV(\hat{f}_1)} \\ \vdots & & \vdots \\ \frac{MRC(r_1, \hat{f}_K)}{PRV(\hat{f}_K)} & \cdots & \frac{MRC(r_p, \hat{f}_K)}{PRV(\hat{f}_K)} \end{pmatrix} + \begin{pmatrix} PRV(\hat{\epsilon}_1) & & \\ & \ddots & \\ & & PRV(\hat{\epsilon}_p) \end{pmatrix} \end{aligned}$$

Then, $\forall i, j = 1, ..., p$

$$\widehat{\Sigma}_{ij} = \sum_{k=1}^{K} \frac{MRC(r_i, \hat{f}_k).MRC(r_j, \hat{f}_k)}{PRV(\hat{f}_k)}; \widehat{\Sigma}_{ii} = \sum_{k=1}^{K} \frac{MRC(r_i, \hat{f}_k)^2}{PRV(\hat{f}_k)} + PRV(\hat{\epsilon}_i) \qquad (25)$$

with PRV and MRC defined by (4) and (7).

4 Sampling properties and simulations under synchronicity

4.1 Sampling properties

In this section, we present the impact of the dimensionality on the estimation errors of $\widehat{\Sigma}$. We are also interested on a comparison with the more popular estimators of the covolatility matrix, namely the adjusted Modulated Realized Covariance estimator MRC^{δ} of Kim Christensen et Al. (2010), the multidimensional kernel estimator $MRker$ of Barndorff-Nielsen et Al. (2008) and the composite realized kernel $\hat{\Sigma}_{comp}$ of LUNDE et Al.(2011). While MRC^{δ} and $MRker$ are defined respectively by (6) and (8), $\hat{\Sigma}_{comp}$ is defined in section 2. The price data are assumed to be synchronized.

The first lemma presents the rate of convergence of the estimator of the integrated volatility of each rotated factor.

Lemma 4.1 *(Rate of convergence of the estimator of the integrated volatility of rotated factors)*
Under the model (9)-(14) and assumption (1)-(2), we have

$$\left|\hat{\Sigma}^{\tilde{f}}_{kk} - \Sigma^{\tilde{f}}_{kk}\right| = O_p\left(n^{-1/4}\right)$$

Secondly, we establish the rate of convergence of estimators of factor loadings vectors $\hat{b}_k = (\hat{b}_{1k}, ..., \hat{b}_{pk})'$, for $k = 1, ..., K$.

Lemma 4.2 *(Rate of convergence of the loading factors)*
Under the model (9)-(14) and assumption (1)-(2), the loading factors satisfy

$$\left\|\hat{b}_k - b_k\right\|_F = \left\|\hat{b}_k - b_k\right\|_2 = O_p\left(p^{1/2} n^{-1/4}\right), \forall\ i = 1, ..., p$$

The following result concerns the rate of convergence of the estimator of the integrated covolatility matrix of the idiosyncratic error term.

Lemma 4.3 *(Rate of convergence of the estimator of the integrated covolatility matrix of the idiosyncratic error term)*
Under the model (9)-(14) and assumption (1)-(2), and under the Frobenius norm

$$\left\|\hat{\Sigma}^{\epsilon} - \Sigma^{\epsilon}\right\|_F = O_p(p^{1/2}n^{-1/4})$$

Having in hand the convergence rate of estimators of loadings, integrated volatility of rotation of factors, integrated covolatility matrix of the idiosyncratic error term, we derive the rate of convergence of the covolatility matrix estimator $\hat{\Sigma}$ and the rates of its competitors.

Theorem 4.1 *(Rate of convergence of the covolatility matrix estimators)*
Under the model (9)-(14) and assumption (1)-(2), we have
i) $\left\|\hat{\Sigma} - \Sigma\right\|_F = O_p(pKn^{-1/4})$
ii) $\left\|MRC^{\delta} - \Sigma\right\|_F = O_p(pn^{-1/5})$
iii) $\left\|MRker - \Sigma\right\|_F = O_p(pn^{-1/5})$
iv) $\left\|\hat{\Sigma}_{comp} - \Sigma\right\|_F = O_p(\sqrt{p(p-1)}n^{-1/5})$ *where* $\|.\|_F$ *is the Frobenius norm defined for a matrix* $A = (a_{ij})_{1\leq i,j\leq p}$ *by* $\|A\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}|a_{ij}|^2}$, MRC^{δ}, $MRker$ *and* $\hat{\Sigma}_{comp}$ *are respectively the adjusted Modulated Realized Covariance, the multidimensional realized kernel estimators and the composite realized kernel.*

From the *Theorem* 4.1, it appears that under the Frobenius norm, the dimensionality of the covolatility matrix reduces the speed of convergence by an order of p. This result is true for all the considered estimators. However, the speed of convergence of $\hat{\Sigma}$ is greater than the one of MRC^{δ}, $MRKer$ or $\hat{\Sigma}_{comp}$. Hence, $\hat{\Sigma}$ has a better performance than MRC^{δ}, $MRker$ and $\hat{\Sigma}_{comp}$ in the estimation of the covolatility matrix Σ. This result will be confirmed using a simulation exercise.

4.2 Simulation under synchronicity

4.2.1 Simulation design

We use a simulation exercise to illustrate and to confirm the finite sample properties and the comparison of our estimators. In this first setting, the simulated prices are synchronized. We take into account the structure of the benchmark model and the identification assumptions.

- The loading factors b is generated such that elements of the k^{th} column $\underline{b}_k$, for $k = 1, ..., K$, follow a normal law with mean 0 and standard deviation 1: $\underline{b}_{ik} \sim N(0, 1)$, $\forall\ i = 1, ..., p$.

- The two factor components in the frictionless return representation are generated by the following model [5]

  – Factor 1

  $$f_{1t} = \sigma_{f1t} dB_{1t}$$

  with B_{1t} a brownian motion and σ_{f1t} generated by a $GARCH$ diffusion model as in Torben G. Andersen et al. (2006)

  $$d\sigma^2_{f1t} = \kappa_{f1} \left(\theta_{f1} - \sigma^2_{f1t} \right) dt + \lambda_{f1} \sigma^2_{f1t} dW_{1t}$$

  with $Corr(W_{1t}, B_{1t}) = -0.5$, $\kappa_{f1} = 0.035$, $\theta_{f1} = 0.636$, $\phi_{f1} = 0.296$, $\lambda_{f1} = \sqrt{2\kappa_{f1}\phi_{f1}}$, $\sigma_{f10} = \theta_{f1}$

  – Factor 2

  $$f_{2t} = \sigma_{f2t} dB_{2t}$$

  with B_{2t} a brownian motion and σ_{f2t} generated by a $GARCH$ diffusion model as in Torben G. Andersen et al. (2006)

  $$d\sigma^2_{f2t} = \kappa_{f2} \left(\theta_{f2} - \sigma^2_{f2t} \right) dt + \lambda_{f2} \sigma^2_{f2t} dW_{2t}$$

  with $Corr(W_{2t}, B_{2t}) = -0.5$, $\kappa_{f2} = 0.035$, $\theta_{f2} = 0.3$, $\phi_{f2} = 0.296$, $\lambda_{f2} = \sqrt{2\kappa_{f2}\phi_{f2}}$, $\sigma_{f20} = \theta_{f2}$

- The idiosyncratic error term in the factor representation is assumed to satisfy

  $$\varepsilon_{it} = \sigma_{it} dW^{\varepsilon}_{it}$$

  with W^{ε}_{it} a brownian motion such that $W^{\varepsilon}_{it} \perp W_{1t}, W_{2t}$ and $W^{\varepsilon}_{it} \perp B_{1t}, B_{2t}$, with the spot volatility generated by three different representative models [6]:

[5] remember that f_{kt} is assumed to be the return of some portfolio

[6] The idea is to simulate the return processes as representative as possible of the literature

– For $1 \leq i \leq p/3$, the volatility of the idiosyncratic component is generated by a Nelson GARCH diffusion limit model as in Bandorff-Nielsen and Shephard (2002):

$$d(\sigma_{it}^2) = (0.1 - \sigma_{it}^2)\, dt + 0.2\sigma_{it}^2 dB_{it}^{\varepsilon}$$

with $Corr(W_{it}^{\varepsilon},\, B_{it}^{\varepsilon}) = -0.3$ and $B_{it}^{\varepsilon} \perp W_{1t}, W_{2t}$ and $B_{it}^{\varepsilon} \perp B_{1t}, B_{2t}$;

– For $p/3 < i \leq 2p/3$, the volatility process is assumed to follow a geometric Ornstein-Uhlenbeck (OU) model as in Bandorff-Nielsen and Shephard (2002):

$$dlog(\sigma_{it}^2) = -0.6\,(0.157 + log(\sigma_{it}^2))\, dt + 0.25 dB_{it}^{\varepsilon}$$

with $Corr(W_{it}^{\varepsilon},\, B_{it}^{\varepsilon}) = -0.3$ and $B_{it}^{\varepsilon} \perp W_t$ and $B_{it}^{\varepsilon} \perp B_t$;

– For $2p/3 < i \leq p$, the volatility follows a $GARCH$ diffusion model as in Andersen et al. (2006):

$$d\sigma_{it}^2 = \kappa_{\varepsilon}\,(\theta_{\varepsilon} - \sigma_{it}^2)\, dt + \gamma_{\varepsilon}\sigma_{it} dB_{it}^{\varepsilon}$$

with $Corr(W_{it}^{\varepsilon},\, B_{it}^{\varepsilon}) = -0.3$ and $B_{it}^{\varepsilon} \perp W_t$ and $B_{it}^{\varepsilon} \perp B_t$; $\kappa_{\varepsilon} = 0.035$, $\theta_{\varepsilon} = 0.636$, $\gamma_{\varepsilon} = 0.296$, $\sigma_{i0} = \theta_{\varepsilon}$

- The slope in the factor representation of the microstructure noise is such that: $c_i \sim N(1, 1)$, $\forall i = 1, ..., p$;

- As in Barndorff-Nielsen et al. (2008), the variance of the microstructure noise of the asset i satisfies the equality: $Var(u_i) = \xi^2 \sqrt{\frac{1}{n}\sum_{t=1}^{n}\sigma_{it}^4}$, with ξ^2 the noise-to-signal ratio which takes values in $\{0.001, 0.005, 0.01\}$ and σ_{it} the spot volatility of the true price process of asset i at time t.

- The variance of the idiosyncratic component η_{it} in the factor representation of the microstructure noise is assumed to have a fraction $1/n^{1.1}$ of the total variance $Var(u_i)$. Then, the variance of the factor term in this representation is given by: $\sigma_g^2 = \frac{\left(\overline{Var(u)} - \overline{\sigma_{\eta}^2}\right)}{\bar{C_p^2}}$, with $\bar{C_p^2} = \frac{1}{p}\sum_{i=1}^{p} c_i^2$.

- g_t and η_{it} are such that: $g_t \sim N(0, \sigma_g^2)$ and $\eta_{it} \sim N(0, \frac{1}{n^{1.1}}Var(u_i))$.

To evaluate performances of our estimators, we firstly simulate the true covolatility matrix using the previous design and it true formula ($\Sigma = bDiag\left[\int_0^1 \sigma_{f1u}^2 du, ..., \int_0^1 \sigma_{fKu}^2 du\right] b' +$

$Diag\left[\int_0^1 \sigma^2_{\varepsilon 1u}du, ..., \int_0^1 \sigma^2_{\varepsilon pu}du\right]$). Secondly, we generate a $p-dimensional$ price vector of noisy prices by adding to the true vectors of prices the microstructure noises as described in (13). Each path of the noisy price vector is constituted of $n+1$ observations. We assume that assets are recorded in a synchronous way, with one trade every 5 minutes for a trading day of 6.5 hours. This corresponds to a number of recorded prices of $n+1=79$ per asset. This number of observations is closed to the one advocated by Lunde et al. (2011). They obtained that in 473 liquid stocks that appeared in the $SP500$, the number of observations per day after synchronization is around 100. We consider three panels, each of them corresponding to a particular level of noise: *panel A*, low noise ($\xi^2 = 0.001$); *panel B*, medium noise ($\xi^2 = 0.005$) and *panel C*, high noise ($\xi^2 = 0.01$). In each panel, we allow the number of assets to take the values $p = 50$ and $p = 100$. Using each path of the noisy vector of prices, we compute the estimate of the covolatility matrix using four competing estimators: our factorial estimator $\hat{\Sigma}$, the composite realized kernel of LUNDE et Al.(2011) ($\hat{\Sigma}_{comp}$), the multivariate kernel ($MRker$), and the adjusted Modulated Realized Covariance estimator (MRC^{δ}).

We run a Monte Carlo exercise to evaluate the average level of errors of these estimators. Errors computed concern the estimation of the integrated covolatility matrix (Columns *Covariance* in the tables), the estimation of the integrated correlation matrix (Columns *Correlation*), and the estimation of the inverse of the integrated covariance matrix (Columns *Inverse*). More specifically, we run 1000 Monte Carlo replications, and for each replication, we compute two statistical criteria to compare estimators. The first is the scaled Frobenius norm of the difference between the estimate and the true matrix. The scaled Frobenius norm is defined by

$$\|A\|_F = \sqrt{\frac{1}{p}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}\left|a_{ij}^2\right|\right)} \tag{26}$$

with $A = (a_{ij})_{i,j=1,...,p}$. As in Hautsh et al.(2009), the idea of scaling our Frobenius norm is to allow comparability as the number of assets increases.
The second criterion is a measure of the positive definiteness, because we also care about the invertibility and the well-conditioning of the estimates. We use the same criterion as in Hautsh et al.(2009)

$$PSD = \begin{cases} 1 \text{ if } \hat{\lambda}_{min} > 0 \\ 0 \text{ otherwise} \end{cases}$$

where $\hat{\lambda}_{min}$ is the smallest estimated eigenvalue of the estimate matrix. In the tables, the columns PSD represent the fraction of integrated covariance estimates that are positive definite in all the Monte Carlo replications.

4.2.2 Simulation results

Tables 1, 2 and 3 present results of the previous simulation design. In 1000 replications, medians of the normalized Fobrenius norms are computed: i) for the estimation errors of the integrated covolatility matrix (Columns *Covariance*); ii) the estimation errors of the integrated correlation matrix (Columns *Correlation*); and iii) the estimation of the inverse of the integrated covariance matrix (Columns *Inverse*). The standard deviations are provided in parenthesis.

We observe that, under the Frobenius norm, for all the considered microstructure noise levels, $\hat{\Sigma}$ outperforms $\hat{\Sigma}_{comp}$, $MRker$ and MRC^{δ} in the estimation of the covolatility matrix. For the four estimators, the estimation errors increase with the dimensionality of the covolatility matrix and with the level of the microstructure noise. These results are consistent with the *theorem* 4.1.

From columns 6 and 7 of tables 1, 2 and 3 (*Diag* and *Off-Diag*) it appears that $\hat{\Sigma}$ dominates as well as in the estimation of integrated volatilities of assets (diagonal terms) than in the estimation of couples of integrated covolatilities (off-diagonal terms).

We also compare through simulations, the four competitors in the estimation of the correlation and the inverse. For these two last statistics, under the Frobenius norm, $\hat{\Sigma}$ is more efficient than $\hat{\Sigma}_{comp}$, $MRker$ and MRC^{δ}. For the special case of the inverse of the covolatility matrix, $\hat{\Sigma}^{-1}$ exists even if the number of assets is bigger than the number of observations per asset (The indicator PSD is always equal to 1 for $\hat{\Sigma}$). But this is not the case for all the other competitors. It is better conditioned than $\hat{\Sigma}_{comp}^{-1}$, $MRker^{-1}$ and $\left(MRC^{\delta}\right)^{-1}$ when the two last matrices exist. These features are the main advantages of $\hat{\Sigma}$ relatively to $MRker$ and MRC^{δ}.

Table 1. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: Low noise

Panel A: Low noise ($\xi^2 = 0.001$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	1.206	0.921	2.271	1	5.183	75.007
	(0.283)	*(0.190)*	*(5.389)*			
$\hat{\Sigma}_{comp}$	1.541	1.055	3.170	1	11.940	119.56338
	(0.292)	*(0.131)*	*(0.197)*			
MRker	1.587	1.128	270.884	1	6.281	134.326
	(0.279)	*(0.129)*	*(74.929)*			
MRC^{δ}	1.560	1.116	147.141	1	6.025	128.28
	(0.279)	*(0.132)*	*(395.209)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	2.310	1.221	2.895	1	18.498	538.322
	(0.485)	*(0.205)*	*(34.790)*			
$\hat{\Sigma}_{comp}$	2.730	1.418	3.835	1	62.401	910.124
	(0.944)	*(0.186)*	*(0.213)*			
MRker	2.778	1.498	NA	0	27.7020	963.417
	(0.923)	*(0.184)*	NA			
MRC^{δ}	2.698	1.480	NA	0	24.707	891.115
	(0.864)	*(0.190)*	*NA*			

Table 2. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: Medium noise

Panel B: Medium noise ($\xi^2 = 0.005$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	2.054	1.228	3.291	1.00	11.823	236.636
	(0.535)	*(0.269)*	*(2.528e+01)*			
$\hat{\Sigma}_{comp}$	2.497	1.511	3.877	1	29.599	356.402
	(0.701)	*(0.188)*	*(3.548e-01)*			
MRker	2.528	1.565	5430.955	0.21	17.353	375.480
	(0.687)	*(0.186)*	*(4.631e+04)*			
MRC^{δ}	2.537	1.588	1035.402	1	16.943	358.348
	(0.571)	*(0.172)*	*(2.682e+03)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	2.876	1.670	6.694	1	26.330	876.362
	(0.775)	*(0.358)*	*(26.258)*			
$\hat{\Sigma}_{comp}$	3.671	2.055	4.116	1	59.837	1489.955
	(0.833)	*(0.272)*	*(0.485)*			
MRker	3.713	2.114	NA	0	38.859	1545.652
	(0.819)	*(0.270)*	NA			
MRC^{δ}	3.634	2.166	NA	0	36.699	1469.797
	(0.711)	*(0.259)*	NA			

Table 3. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: High noise

Panel C: High noise ($\xi^2 = 0.01$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	2.237	1.399	5.285	1.00	13.964	266.213
	(0.559)	*(0.260)*	*(4.066e+01)*			
$\hat{\Sigma}_{comp}$	2.743	1.693	4.200	1	28.800	409.016
	(0.599)	*(0.176)*	*(8.303e-01)*			
MRker	2.783	1.733	7283.756	0.88	19.653	427.489
	(0.588)	*(0.172)*	*(3.667e+06)*			
MRC^{δ}	2.832	1.749	5054.455	1	20.112	431.136
	(0.575)	*(0.170)*	*(5.737e+03)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	3.441	1.852	6.373	1	35.335	1356.935
	(0.974)	*(0.457)*	*(31.752)*			
$\hat{\Sigma}_{comp}$	4.687	2.282	4.829	1	72.978	2230.337
	(0.973)	*(0.299)*	*(0.863)*			
MRker	4.737	2.325	NA	0	52.645	2291.703
	(0.962)	*(0.295)*	*NA*			
MRC^{δ}	4.543	2.335	NA	0	52.526	2208.819
	(0.883)	*(0.279)*	*NA*			

5 The asynchronous case

5.1 The estimator

In this part we make the realistic assumption that the asset prices are not synchronous, and transactions occur in a random and asynchounous way. We propose to keep the same estimator as in the synhronous case. It means $\hat{\Sigma}$ defined as in (25) by: $\forall i, j = 1, ..., p$

$$\widehat{\Sigma}_{ij} = \sum_{k=1}^{K} \frac{MRC(r_i, \hat{f}_k).MRC(r_j, \hat{f}_k)}{PRV(\hat{f}_k)}; \widehat{\Sigma}_{ii} = \sum_{k=1}^{K} \frac{MRC(r_i, \hat{f}_k)^2}{PRV(\hat{f}_k)} + PRV(\hat{\epsilon}_i) \qquad (27)$$

with PRV and MRC defined by (4) and (7)

Comparing to the multidimensional kernel and to the adjusted Modulated Realized Covariance, the loss of observations through the synchronization process is less important with $\widehat{\Sigma}$. Indeed, the computation of $\widehat{\Sigma}$ need only to synchronize two series: p_i and $p_{\cdot}$. But with the two others estimators ($MRker$ and MRC^{δ}), the synchronization process is applied to p series of prices. This leads to a lost of more observations than in the $\widehat{\Sigma}$ case. From this observation, and using the fact that $\widehat{\Sigma}$ uses the structure of the model (the others estimators don't do it), we guess that $\widehat{\Sigma}$ will outperform $MRker$ and MRC^{δ}.

We use a simulation exercise to confirm the superiority $\widehat{\Sigma}$ relatively to $\hat{\Sigma}_{comp}$, $MRker$ and MRC^{δ} for the asynchonous case.

5.2 Simulations in the asynchronous case

The simulation design in this setting is different from the first one in one main point: the trading time. The set of assets is divided in three groups. In the first group, a trade occurs every 30 seconds. The trading time is generated by a *poisson process* of mean 3 in the second group ($t_i \sim Poisson(3)$). In other words, on average, there is one trade every 90 seconds in the second group of assets. For assets in the group 3, since the trading time is generated by a *poisson process* of mean 5 ($t_i \sim Poisson(5)$), a trade occurs on average every 150 seconds. In all the different groups, price data are not regularly spaced.

The synchronization of the simulated data is done using the refesh time of Barndorff-Nielsen et Al. (2008). We evaluate the performance of our competing estimators using the same statistical criteria as in the synchronous case: the Frobenius norm of the estimation error and the fraction of integrated covariance estimates positive definite in all the Monte Carlo replications. The tables below give the simulation results.

Comments concerning the simulation results are roughly the same than in the synchronous case: independently of the microstructure noise level and under the Frobenius norm, $\hat{\Sigma}$ performs better than $\hat{\Sigma}_{comp}$, $MRker$ and MRC^{δ} in the estimation of the covolatility, the correlation and the inverse matrices. The use of the model structure, and a smaller loss of data can explain this good performance of $\hat{\Sigma}$. $\hat{\Sigma}^{-1}$ always exists, and is well conditioned.

Table 4. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: Low noise

Panel A: Low noise ($\xi^2 = 0.001$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	6.291	1.643	2.877	1	121.494	1897.006
	(0.609)	*(0.121)*	*(0.311)*			
$\hat{\Sigma}_{comp}$	6.571	1.855	3.403	1	117.091	2099.866
	(0.580)	*(0.109)*	*(0.126)*			
MRker	6.808	1.932	215.077	1	117.156	2197.824
	(0.591)	*(0.131)*	*(91.139)*			
MRC^{δ}	6.719	1.907	136.493	1	115.651	2163.327
	(0.582)	*(0.131)*	*(383.858)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	6.056	2.015	3.325	1	161.147	3628.646
	(0.426)	*(0.146)*	*(0.327)*			
$\hat{\Sigma}_{comp}$	6.625	2.234	3.898	1	133.563	4387.458
	(0.374)	*(0.083)*	*(0.147)*			
MRker	6.940	2.400	NA	0	151.158	4761.590
	(0.373)	*(0.094)*	NA			
MRC^{δ}	6.700	2.314	NA	0	139.204	4433.488
	(0.368)	*0(.098)*	*NA*			

Table 5. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: Medium noise

Panel B: Medium noise ($\xi^2 = 0.005$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	5.496	1.872	4.089	1	74.317	1481.721
	(0.506)	*(0.125)*	*(2.725e-01)*			
$\hat{\Sigma}_{comp}$	6.112	1.945	3.555	1	73.014	1819.152
	(0.459)	*(0.103)*	*(1.656e-01)*			
MRker	6.277	2.032	3526.001	0.2	79.613	1928.104
	(0.500)	*(0.116)*	*(1.931e+05)*			
MRC^{δ}	6.087	1.947	490.138	1	79.781	1795.881
	(0.477)	*(0.103)*	*(5.315e+02)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	6.845	2.253	4.188	1	144.688	4609.252
	0.483	*0.228*	*3.023*			
$\hat{\Sigma}_{comp}$	8.001	2.666	3.842	1	145.256	6251.748
	(0.443)	*(0.111)*	*(0.192)*			
MRker	8.307	2.850	NA	0	167.138	6793.077
	(0.465)	*(0.138)*	NA			
MRC^{δ}	(8.164)	(2.783)	NA	0	162.433	6484.103
	(0.438)	*(0.143)*	NA			

Table 6. Performance of the factorial estimator relatively to the composite kernel, the multivariate Kernel and the adjusted modulated realized covariance: High noise

Panel C: High noise ($\xi^2 = 0.01$)						
Number of assets: N=50						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	4.803	1.863	4.135	1	83.121	1100.954
	(0.325)	*(0.103)*	*(3.850e+00)*			
$\hat{\Sigma}_{comp}$	5.578	1.997	3.252	1	88.978	1485.581
	(0.352)	*(0.088)*	*(2.058e-01)*			
MRker	5.740	2.128	5720.101	0.74	94.496	1570.682
	(0.328)	*(0.085)*	*(1.127e+04)*			
MRC^{δ}	5.645	2.053	986.788	1	95.896	1507.651
	(0.331)	*(0.098)*	*(1.530e+03)*			
Number of assets: N=100						
	Covariance	Correlation	Inverse	PSD	Diag	Off-Diag
$\hat{\Sigma}$	7.218	2.552	4.473	1	255.925	5117.892
	(0.551)	*(0.213)*	*(3.002)*			
$\hat{\Sigma}_{comp}$	8.352	3.007	5.350	1	171.577	7003.717
	(0.458)	*(0.093)*	*(0.625)*			
MRker	8.626	3.218	NA	0,000	202.4798	7432.442
	(0.449)	*(0.130)*	*NA*			
MRC^{δ}	8.609	3.162	NA	0	205.509	7255.146
	(0.491)	*(0.143)*	*NA*			

6 Empirical studies

6.1 Data

We based our empirical application on intraday data collected from the TAQ database of the Wharton Research Data Services. Our sampling period goes from January 2007 to December 2011. The original sample contained high frequency data of more than 500 stocks. As in Lunde et al. (2011), we removed stocks traded less than 195 times during a given trading day. In order to clean the database, we used the cleaning procedure of Barndorff-Nielsen et al. (2009) as advocated in the literature. The cleaned data set contain around 384 assets. Since for different stocks trades don't occur at the same time, we need a synchronization process in order to construct the different estimators. The data synchronization is done using the refesh time approach of Barndorff-Nielsen et al. (2008). After the synchonization process applied to all trading days of our database, the synchronized data set contain on average 100 intraday observations.

6.2 The choice of the number of factors in practice

In the above estimation methodology, the number of factors is assumed to be known. But in practice, we need to estimate it. We use the approach developed in Bai et al. (2008) in order to estimate the number of factors. This approach consists on choosing a penalty function $g(.,.)$, and an information criteria IC defined by

$$IC = log\left[\frac{1}{p}\sum_{j=1}^{\lfloor 1/\Delta \rfloor}(r^*_{j\Delta} - bf_{j\Delta})'(r^*_{j\Delta} - bf_{j\Delta})\right] + Kg(p, \lfloor 1/\Delta \rfloor) \tag{28}$$

The number of factors is chosen by looking for the value $\hat{K}$ of K which minimizes IC. We use as penalty function $g(p, \lfloor 1/\Delta \rfloor) = \frac{p+\lfloor 1/\Delta \rfloor}{p\lfloor 1/\Delta \rfloor} \times log\left[\frac{p\lfloor 1/\Delta \rfloor}{p+\lfloor 1/\Delta \rfloor}\right]$, as in Bai et al.(2008). With this approach, the number of factors is not constant over the sampling period, it changes from a trading day to another one. From our data set, we obtain it values vary from 1 to 4, with an average equals to 3.

6.3 Performance evaluation

We want to use recorded data in order to compare our estimators of the covolatility with the other competitors. The comparison approach will be the same as in LUNDE et al.(2015). The idea is to compare the ability of each type of estimator to minimize the risk (portfolio variance) during a portfolio allocation. To be more precise, let's assume that the estimator $\hat{\Sigma}_t$ is computed over the trading day t. $\hat{\Sigma}_t$ is going to be used as the forecast of the covolatility matrix during the day $t+1$, under a random walk forecasting model. Thus, for the trading day $t+1$, we are going firstly to resolve the following classic Markowitz portfolio problem with a gross exposure constraint

$$\begin{cases} Min \quad w'_{t+1}\hat{\Sigma}_t w_{t+1} \\ s.t\ w'_{t+1}1 = 1 \quad and \quad \sum_{i=1}^{N} |w_{i,t+1}| \leq 1 + 2s \end{cases} \tag{29}$$

which leads to a weight vector $\hat{w}_{t+1}$. Here, s represents the share of the stocks in the portfolio which can be held short. We will allow s to take the values in $0, 0.25, 0.5, 1$. While for $s = 0$ the portfolio is completely long, it is totally short for $s = 1$. Adding a gross exposure constraint is useful for our setup since it moderates the error accumulation effect from the covolatility estimation to the Risk approximation when gross exposure s

is moderate (see for example (Jianqing Fan et al. (2012)). Also, with a gross exposure constraint, the Markowitz portfolio problem is guarantee to have a solution even for semi-definite positive matrices. The latter is useful since some of our competitors are just semi-definite positive and not definite positive ($MRker$ and MRC^{δ}).

The performance of a given estimator $\hat{\Sigma}_t$ during the trading day $t+1$ is evaluated as in Lunde et al.(2011) using the following criteria

$$\hat{w}_{t+1}^{'} RCov_{t+1} \hat{w}_{t+1} \tag{30}$$

where $RCov_{t+1}$ is the out of sample *5-minutes* realized covariance matrix.

For a given covolatility matrix estimator and a given gross exposure level s, the performance over the entire sample will be the average over of the daily performance. We add the naive equally weighted portfolio in the set of competitors as advocated by DeMiguel (2009). We also add a second version of our factorial estimator called $\hat{\Sigma}_2$, which is derived as product of the correlation matrix associated to $\hat{\Sigma}$ with diagonal matrices of optimal estimators of integrated volatility of each asset

$$\begin{aligned}\hat{\Sigma}_2 = Diag(PRV(r_1), ..., PRV(r_p)) \times Diag(\hat{\Sigma}_{11}, ..., \hat{\Sigma}_{pp})^{-1} \times \hat{\Sigma} \\ \times Diag(\hat{\Sigma}_{11}, ..., \hat{\Sigma}_{pp})^{-1} \times Diag(PRV(r_1), ..., PRV(r_p))\end{aligned}$$

The following table is a summary of the results we obtained

Table 7. Performance evaluation of estimators of the covolatilaty matrix: liquid assets

	s=0	s=0.25	s=0.5	s=1
$\hat{\Sigma}$	**0.495**	**0.245**	**0.240**	**0.241**
$\hat{\Sigma}_2$	**0.508**	**0.240**	**0.233**	**0.235**
$\hat{\Sigma}_{comp}$	0.520	0.301	0.325	0.326
$MRKer$	0.545	0.278	0.263	0.258
MRC^{δ}	0.559	0.362	0.343	0.319
$EqualWeight$	0.636	0.636	0.636	0.636

From the empirical performance results, it appears that ours factorial estimators ($\hat{\Sigma}$ and $\hat{\Sigma}_2$) work well relatively to the competitors. The dominance is quasi total for all the different gross exposure levels. If the comparison analysis was based on a function of the inverse of the covolatility matrix, the gaps with the competitors would be more important.

Table 8. Performance evaluation of estimators of the covolatilaty matrix: illiquid assets

	s=0	s=0.25	s=0.5	s=1
$\hat{\Sigma}$	**0.332**	**0.267**	**0.264**	**0.263**
$\hat{\Sigma}_2$	**0.332**	**0.253**	**0.253**	**0.253**
$\hat{\Sigma}_{comp}$	0.407	0.377	0.378	0.377
$MRKer$	0.396	0.315	0.312	0.336
MRC^{δ}	0.411	0.337	0.335	0.349
$EqualWeight$	0.481	0.481	0.481	0.481

7 Concluding remarks

This paper presents a factor based estimation methodology of the covolatility matrix when the number of assets is large. The framework assumes a statistical factor representation of both the frictionless returns and the microstructure noise. We take advantage of the properties of the pre-averaging estimator of Jacod et al. (2009) in order to construct our estimator.

We prove analytically that when the observed asset prices are synchronized, independently of the microstructure noise level, our estimator outperforms the more popular estimators of the covolatility matrix, namely the composite kernel of Lunde et al.(2015), the multidimensional *kernel* estimator (of Barndorff-Nielsen et al. (2008)) and the adjusted *Modulated Realized Covariance* estimator (of Christensen et Al. (2010)). This result is confirmed through an intensive simulation exercise. In addition, simulations illustrate the superiority of our estimator in the estimation of the correlation and inverse matrices. Unlike the other estimators, the inverse of our estimator always exists and is well conditioned. The measures of the performance used in the simulation exercise are the normalized Frobenius norm of the estimation error and the fraction of integrated covariance estimates that are positive definite in all the Monte Carlo replications.

Under the non-synchronicity assumption, results obtained are roughly the same than in the synchronous case. A smaller number of data loss comparing to the competitors during the synchronization exercise, is one explanation of the good performance of our estimator.

Using a five years TAQ database of the Wharton Research Data Services, we conduct empirical studies in order to confirm the superiority of our approach. The performance criteria is based on a measure of the risk associated to the Markowitz portfolio under a gross exposure constraint. The result is a quasi total dominance of our approach independently of the gross exposure level.

References

[1] Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. 2003, Modeling and Forecasting Realized Volatility, Econometrica, 71, 529-626.

[2] Barndorff-Nielsen, O.E., P.R. Hansen, Lunde, A. and Shephard, N. (2008a), "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," Econometrica, 76, 1481-1536.

[3] Barndorff-Nielsen, O.E., P.R. Hansen, Lunde, A. and Shephard, N. (2008b), "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading," working paper, Oxford University.

[4] Christensen, K., Kinnebrock, S., and Podolskij, M. (2010), "'Pre-Averaging Estimators of the Ex-Post Covariance Matrix in Noisy Diffusion Models With Non-Synchronous Data", Journal of Econometrics, 159, 116Ű133.

[5] Fan, J., Fan, Y., and Lv, J. (2008), "High Dimensional Covariance Matrix Estimation Using a Factor Model", Journal of Econometrics, 147, 186-197.

[6] Fan, J., Li, Y., and Yu, K. (2012), "Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection", Journal of the American Statistical Association, 107:497, 412-428.

[7] Hayashi, T., and Yoshida, N. (2005), "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes", Bernoulli, 11, 359-379.

[8] Zhang,L., MYKLAND, P., and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data", Journal of the American Statistical Association December 2005, Vol. 100, No. 472.

[9] Hautsch, N., Podolskij, M. (2010), "Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence", SFB 649 Discussion Paper 2010-038.

[10] Hautsch, N., Kyj, L., and Oomen, R., (2009), "A blocking and regularization approach to high dimensional realized covariance estimation", SFB 649 Discussion Paper 2009-049.

[11] Hansen, P., and Lunde, A., (2006), "Realized Variance and Market Microstruture Noise", Journal of Business and Economic Statistics, Vol. 24, No.2.

[12] Podolskij and M. Vetter, (2009), "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps", Bernoulli 15(3), 634-658.

[13] Tao, M., Wang, Y., and Chen, X. (2011), "Fast Convergence Rates in Estimating

Large Volatility Matrices Using High-Frequency Financial Data", Journal of the American Statistical Association, September 2011, Vol. 106, No. 495.
[14] Andersen, T., Bollerslev, T., and Meddahi, N., (2011), "Realized Volatility Forecasting and Market Microstructure Noise", Journal of Econometrics, 160, 220-234.
[15] Wang, Y., and Zou, J. (2010), "Vast Volatility Matrix Estimation for High-Frequency Financial Data", The Annals of Statistics, 38, 943-978.
[16] Wang, Y., and Zou, J. (2010), "Estimating Covariation: Epps Effect, Microstructure Noise", Journal of Econometrics, 160, 33-47.
[17] Zheng, X., and Li, Y. (2011), "On the estimation of integrated covariance matrices of high dimensional diffusion processes", The Annals of Statistics, Vol. 39, No.6, 3121-3151.

Appendix: Technical Proofs

Proof of Theorem 3.1

A) Proof of: $\frac{1}{p}S\hat{W}'bf_t \overset{p}{\rightarrow} f_t$

This proof consists on 12 steps inspired from the paper of Stock and Watson (2002).

Step 1: $\frac{1}{p}\sum_{i=1}^{p}\epsilon_{it}^2 \sim O_p(1)$

We assume that:

- *A1)* $\underset{p\longrightarrow\infty}{lim}\ \underset{t}{Sup}\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|E(\epsilon_{it}\epsilon_{jt})| < \infty$;
- *A2)* $\underset{p\longrightarrow\infty}{lim}\ \underset{t,s}{Sup}\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|Cov(\epsilon_{is}\epsilon_{it},\epsilon_{js}\epsilon_{jt})| < \infty$

Since

$$\frac{1}{p}\sum_{i=1}^{p}\epsilon_{it}^2 = \frac{1}{p}\sum_{i=1}^{p}E\left[\epsilon_{it}^2\right] + \frac{1}{p}\sum_{i=1}^{p}\left[\epsilon_{it}^2 - E\left(\epsilon_{it}^2\right)\right]$$

we just need to prove that

$$\frac{1}{p}\sum_{i=1}^{p}E\left[\epsilon_{it}^2\right] \sim O(1) \text{ and } \frac{1}{p}\sum_{i=1}^{p}\left[\epsilon_{it}^2 - E\left(\epsilon_{it}^2\right)\right] \sim o_p(1)$$

The following inequalities hold:

$$\frac{1}{p}\sum_{i=1}^{p}E\left[\epsilon_{it}^2\right] \leq \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|E(\epsilon_{it}\epsilon_{jt})| \leq \underset{t}{Sup}\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|E(\epsilon_{it}\epsilon_{jt})|$$

Since $\underset{t}{Sup}\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|E(\epsilon_{it}\epsilon_{jt})|$ converges, it is bounded. Thus $\frac{1}{p}\sum_{i=1}^{p}E\left[\epsilon_{it}^2\right]$ is bounded, it means $O(1)$. In addition

$$\begin{aligned} E\left[\left(\frac{1}{p}\sum_{i=1}^{p}\left[\epsilon_{it}^2 - E\left(\epsilon_{it}^2\right)\right]\right)^2\right] &= \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}E\left[(\epsilon_{it}^2 - E(\epsilon_{it}^2))(\epsilon_{jt}^2 - E(\epsilon_{jt}^2))\right] \\ &= \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}Cov\left(\epsilon_{it}^2,\epsilon_{jt}^2\right) \\ &\leq \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\left|Cov\left(\epsilon_{it}^2,\epsilon_{jt}^2\right)\right| \\ &\leq \underset{t,s}{Sup}\ \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\left|Cov\left(\epsilon_{it}\epsilon_{is},\epsilon_{jt}\epsilon_{js}\right)\right| \end{aligned}$$

Since $\underset{t,s}{Sup}\ \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\left|Cov\left(\epsilon_{it}\epsilon_{is},\epsilon_{jt}\epsilon_{js}\right)\right|$ is bounded, it follows that

$$\underset{t,s}{Sup}\ \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\left|Cov\left(\epsilon_{it}\epsilon_{is},\epsilon_{jt}\epsilon_{js}\right)\right| \longrightarrow 0$$

We deduce that $\frac{1}{p}\sum_{i=1}^{p}\left[\epsilon_{it}^2 - E\left(\epsilon_{it}^2\right)\right]$ converges in $2-mean$ to 0. Hence

$$\frac{1}{p}\sum_{i=1}^{p}\left[\epsilon_{it}^2 - E\left(\epsilon_{it}^2\right)\right] \xrightarrow{p} 0$$

Step 2: Let $\Gamma = \{\gamma \in \Re^p / \gamma'\gamma/p = 1\}$. We want to prove that $\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p^2}\gamma' IV(\epsilon)\gamma \longrightarrow 0$ as $p \longrightarrow \infty$, with $IV(\epsilon) = Diag(IV(\epsilon_1), ..., IV(\epsilon_p))$. We make the additional assumption that $\forall\ \ i = 1, ..., p$, the quadratic variation of the idiosyncratic component ϵ_i is bounded by a scalar M. Thus, we can write

$$\begin{aligned}
\frac{1}{p^2}\gamma' IV(\epsilon)\gamma &= \frac{1}{p^2}\sum_{i=1}^{p}\gamma_i^2 . IV(\epsilon_i)\\
&\leq \left[\frac{1}{p^2}\sum_{i=1}^{p}\gamma_i^4\right]^{1/2}\left[\frac{1}{p^2}\sum_{i=1}^{p} IV(\epsilon_i)^2\right]^{1/2}\\
&\leq \left[\frac{1}{p^2}\left(\sum_{i=1}^{p}\gamma_i^2\right)^2\right]^{1/2}\left[\frac{1}{p^2}\sum_{i=1}^{p} IV(\epsilon_i)^2\right]^{1/2}\\
&\leq \left(\frac{1}{p}\gamma'\gamma\right).\left[\frac{1}{p^2}\sum_{i=1}^{p} IV(\epsilon_i)^2\right]^{1/2}\\
&\leq \left[\frac{1}{p^2}\sum_{i=1}^{p} IV(\epsilon_i)^2\right]^{1/2}\\
&\leq \left(\frac{M^2}{p}\right)^{1/2}
\end{aligned}$$

We deduce that $\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p^2}\gamma' IV(\epsilon)\gamma \longrightarrow 0$ as $p \longrightarrow \infty$.

Step 3: If $\int_0^1 E(q_t^2)dt \sim O(1)$ then $\underset{\gamma\in\Gamma}{Sup}\left|\frac{1}{p}\int_0^1 E(q_t.\gamma'\epsilon_t)dt\right| \longrightarrow 0$ as $p \longrightarrow \infty$

$$\begin{aligned}
\frac{1}{p}\int_0^1 E(q_t.\gamma'\epsilon_t)dt &\leq \int_0^1\left[E(q_t^2)\right]^{1/2}\left[E\left(\left(\frac{1}{p}\gamma'\epsilon_t\right)^2\right)\right]^{1/2}dt\\
&\leq \left[\int_0^1 E(q_t^2)dt\right]^{1/2}.\left[\int_0^1 E\left(\left(\frac{1}{p}\gamma'\epsilon_t\right)^2\right)dt\right]^{1/2}\\
&\leq O(1).\left[\frac{1}{p^2}\int_0^1\gamma' E(\epsilon_t\epsilon_t')\gamma dt\right]^{1/2}\\
&\leq O(1).\left[\frac{1}{p^2}\gamma' IV(\epsilon)\gamma\right]^{1/2}
\end{aligned}$$

The first inequality comes from the Cauchy-Schwarz inequality and the second from the Holder inequality. From the last inequality, the result is deduced using the *step 2*.

Step 4: $\underset{\gamma\in\Gamma}{Sup}\left|\frac{1}{p}\int_0^1 E(f_{kt}.\gamma'\epsilon_t)dt\right| \longrightarrow 0$ as $p \longrightarrow 0$, $\forall k = 1, ..., K$

This result is obtained from the *step 3* by taking $q_t = f_{kt}$. Indeed, $\int_0^1 E(f_{kt}^2)dt = IV(f_k) < \infty$ by assumption.

Step 5: Assume that *A3)* $\frac{b'b}{p} \longrightarrow I_K$ as $p \longrightarrow 0$. Then $\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p}\gamma' b\int_0^1 E(f_t.\gamma'\epsilon_t)dt \longrightarrow 0$ as $p \longrightarrow 0$

$$
\begin{aligned}
\frac{1}{p}\gamma' b \int_0^1 E(f_t.\gamma'\epsilon_t)dt &= \sum_{k=1}^{K} \gamma' \frac{b_k}{p} . \int_0^1 E(f_{kt}.\frac{\gamma'}{p}\epsilon_t)dt \\
&= \sum_{k=1}^{K} \left(\gamma' \frac{b_k}{p}\right) . \int_0^1 E\left(f_{kt}.\left(\frac{1}{p}\sum_{i=1}^{p}\gamma_i\epsilon_{it}\right)\right)dt \\
&\leq \sum_{k=1}^{K} \left|\gamma' \frac{b_k}{p}\right| . \left|\int_0^1 E\left(f_{kt}.\left(\frac{1}{p}\sum_{i=1}^{p}\gamma_i\epsilon_{it}\right)\right)dt\right| \\
\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p}\gamma' b \int_0^1 E(f_t.\gamma'\epsilon_t)dt &\leq \left[\underset{k}{Max}\, \underset{\gamma\in\Gamma}{Sup} \left|\gamma' \frac{b_k}{p}\right|\right] . \sum_{k=1}^{K} \underset{\gamma\in\Gamma}{Sup} \left|\int_0^1 E\left(f_{kt}.\left(\frac{1}{p}\sum_{i=1}^{p}\gamma_i\epsilon_{it}\right)\right)dt\right| \\
&\leq \left\{\underset{\gamma\in\Gamma}{Sup}\, (\gamma'\gamma/p)^{1/2}\right\} . \left\{\underset{k}{Max}\, (\underline{b}'_k \underline{b}_k/p)^{1/2}\right\} \\
&\quad \times \sum_{k=1}^{K} \underset{\gamma\in\Gamma}{Sup} \left|\int_0^1 E\left(f_{kt}.\left(\frac{1}{p}\sum_{i=1}^{p}\gamma_i\epsilon_{it}\right)\right)dt\right|
\end{aligned}
$$

From the definition of Γ and assumption *A3)*, as $p \longrightarrow \infty$,

$$\underset{\gamma\in\Gamma}{Sup}\, (\gamma'\gamma/p)^{1/2} \longrightarrow 1 \text{ and } \underset{k}{Max}\, (\underline{b}'_k \underline{b}_k/p)^{1/2} \longrightarrow 1$$

In addition, from *step 4*, as $p \longrightarrow \infty$,

$$\sum_{k=1}^{K} \underset{\gamma\in\Gamma}{Sup} \left|\int_0^1 E\left(f_{kt}.\left(\frac{1}{p}\sum_{i=1}^{p}\gamma_i\epsilon_{it}\right)\right)dt\right| \longrightarrow 0$$

Then $\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p}\gamma' b \int_0^1 E(f_t.\gamma'\epsilon_t)dt \longrightarrow 0$ as $p \longrightarrow \infty$.

Step 6*:* Define $\forall\gamma \in \Gamma$, $R(\gamma) = \frac{1}{p^2}\gamma'\Sigma\gamma$ and $R^*(\gamma) = \frac{1}{p^2}\gamma' b.IV(f).b'\gamma$. Then $\underset{\gamma\in\Gamma}{Sup}\, |R(\gamma) - R^*(\gamma)| \longrightarrow 0$ as $p \longrightarrow \infty$.

$$
\begin{aligned}
|R(\gamma) - R^*(\gamma)| &= \left|\frac{1}{p^2}\gamma'\Sigma\gamma - \frac{1}{p^2}\gamma' b.IV(f).b'\gamma\right| \\
&= \left|\frac{1}{p^2}\gamma' \left[b.IV(f).b' + IV(\epsilon)\right]\gamma - \frac{1}{p^2}\gamma' bIV(f)b'\gamma\right| \\
&= \left|\frac{1}{p^2}\gamma' IV(\epsilon)\gamma\right| \\
&\leq \underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p^2}\gamma' IV(\epsilon)\gamma
\end{aligned}
$$

Hence, $\underset{\gamma\in\Gamma}{Sup}\, |R(\gamma) - R^*(\gamma)| \leq \underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p^2}\gamma' IV(\epsilon)\gamma$. Since $\underset{\gamma\in\Gamma}{Sup}\, \frac{1}{p^2}\gamma' IV(\epsilon)\gamma \longrightarrow 0$ as $p \longrightarrow \infty$ by the *step 2*, we deduce that $\underset{\gamma\in\Gamma}{Sup}\, |R(\gamma) - R^*(\gamma)| \longrightarrow 0$ as $p \longrightarrow \infty$.

Step 7*:* $\left|\underset{\gamma\in\Gamma}{Sup}\, R(\gamma) - \underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma)\right| \longrightarrow 0$ as $p \longrightarrow \infty$

From the properties of the *Sup*

$$\left|\underset{\gamma\in\Gamma}{Sup}\, R(\gamma) - \underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma)\right| \leq \underset{\gamma\in\Gamma}{Sup}\, |R(\gamma) - R^*(\gamma)|$$

Since $\underset{\gamma\in\Gamma}{Sup}\, |R(\gamma) - R^*(\gamma)| \longrightarrow 0$ as $p \longrightarrow \infty$ from the *step 7*, the result is obtained.

Step 8: $\underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma) \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$, with $IV(f)_{11}$ the element in the first line and first column of $IV(f)$

We consider the following Choleski decomposition of $b'b/p$

$$\frac{b'b}{p} = \left(\frac{b'b}{p}\right)^{1/2} \left(\frac{b'b}{p}\right)^{1/2'}$$

There exist two vectors δ and V such that γ can be represented in the following way

$$\gamma = b\,(b'b/p)^{-1/2}\,\delta + V, \text{ with } V'b = 0 \text{ and } \delta'\delta \leq 1$$

From the previous specification, we derive the following expression of $R^*(\gamma)$

$$\begin{aligned}
R^*(\gamma) &= \tfrac{1}{p^2}\gamma' b IV(f) b'\gamma \\
&= \tfrac{1}{p^2}\left[b\left(\tfrac{b'b}{p}\right)^{-1/2}\delta + V\right]' b.IV(f).b' \left[b\left(\tfrac{b'b}{p}\right)^{-1/2}\delta + V\right] \\
&= \left[\delta'\left(\tfrac{b'b}{p}\right)^{-1/2} b' + V'\right] \tfrac{b}{p}.IV(f).\tfrac{b'}{p}\left[b\left(\tfrac{b'b}{p}\right)^{-1/2}\delta + V\right] \\
&= \left[\delta'\left(\tfrac{b'b}{p}\right)^{1/2} + \tfrac{V'b}{p}\right].IV(f).\left[\left(\tfrac{b'b}{p}\right)^{1/2}\delta + \tfrac{b'V}{p}\right] \\
&= \delta'\left(\tfrac{b'b}{p}\right)^{1/2}.IV(f).\left(\tfrac{b'b}{p}\right)^{1/2}\delta
\end{aligned}$$

Then,

$$\begin{aligned}
\underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma) &= \underset{\delta,\delta'\delta\leq 1}{Sup}\left\{\delta'\left(\tfrac{b'b}{p}\right)^{1/2}.IV(f).\left(\tfrac{b'b}{p}\right)^{1/2}\delta\right\} \\
&= \textit{Largest eigenvalue of } \left(\tfrac{b'b}{p}\right)^{1/2}.IV(f).\left(\tfrac{b'b}{p}\right)^{1/2} \\
&\equiv \hat{\sigma}_{11}
\end{aligned}$$

Since $\frac{b'b}{p} \longrightarrow I_K$ as $p \longrightarrow \infty$, we have

$$\left(\frac{b'b}{p}\right)^{1/2} IV(f) \left(\frac{b'b}{p}\right)^{1/2} \xrightarrow{p} IV(f)$$

By the continuity of eigenvalues, $\hat{\sigma}_{11} \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$. This leads to $\underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma) \longrightarrow IV(f)_{11}$.

Step 9: $\underset{\gamma\in\Gamma}{Sup}\, R(\gamma) \longrightarrow IV(f)_{11}$

Since $\underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma) \longrightarrow IV(f)_{11}$ from the *step 8*, and since $\left|\underset{\gamma\in\Gamma}{Sup}\, R(\gamma) - \underset{\gamma\in\Gamma}{Sup}\, R^*(\gamma)\right| \longrightarrow 0$ as $p \longrightarrow \infty$ from the *step 7*, we conclude that $\underset{\gamma\in\Gamma}{Sup}\, R(\gamma) \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$.

Step 10: If $\hat{b}_1 = Arg\,\underset{\gamma\in\Gamma}{Sup}\, R(\gamma)$ then $R^*(\hat{b}_1) \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$

If $\hat{b}_1 = Arg\,\underset{\gamma\in\Gamma}{Sup}\, R(\gamma)$, then $R(\hat{b}_1) = \underset{\gamma\in\Gamma}{Sup}\, R(\gamma)$. We derive from the *step 9* that $R(\hat{b}_1) \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$. In addition,

$$\left|R(\hat{b}_1) - R^*(\hat{b}_1)\right| \leq \underset{\gamma\in\Gamma}{Sup}\left|R(\gamma) - R^*(\gamma)\right| \longrightarrow 0 \; as \; p \longrightarrow \infty$$

Hence, $\left|R(\hat{b}_1) - R^*(\hat{b}_1)\right| \longrightarrow 0$ as $p \longrightarrow \infty$. This latter result together with $R(\hat{b}_1) \longrightarrow IV(f)_{11}$ leads to $R^*(\hat{b}_1) \longrightarrow IV(f)_{11}$ as $p \longrightarrow \infty$.

Step 11: Let $\underline{W}_1$ denotes the first column of W (the matrix of ordered eigenvectors of Σ). $\underline{W}_1$ is the eigenvector of Σ associated to its largest eigenvalue. We also define the variable S_1 by: $S_1 = 1$ if $\underline{W'}_1\underline{b}_1 \geq 0$ and $S_1 = -1$ if $\underline{W'}_1\underline{b}_1 \leq 0$, with $\underline{b}_1$ the first column of the loading matrix b. Then $S_1\frac{\underline{W'}_1 b}{p} \xrightarrow{p} l_1'$, with $l_1 = (1, 0, ..., 0)'$.
There exist $\hat{\delta}$ and $\hat{V}$ such that $\underline{W}_1 = b\left(\frac{b'b}{p}\right)^{-1/2}\hat{\delta} + \hat{V}$, with $\hat{V}'b = 0$ and $\hat{\delta}'\hat{\delta} \leq 1$. Let's take $C_{NT} = \left(\frac{b'b}{p}\right)^{1/2}.IV(f).\left(\frac{b'b}{p}\right)^{1/2}$. It follows that $R^*(\underline{W}_1) = \hat{\delta}'.C_{NT}.\hat{\delta}$. Thus

$$\begin{aligned} R^*(\underline{W}_1) - IV(f)_{11} &= \hat{\delta}'\left(C_{NT} - IV(f)\right)\hat{\delta} + \hat{\delta}'.IV(f).\hat{\delta} - IV(f)_{11} \\ &= \hat{\delta}'\left(C_{NT} - IV(f)\right)\hat{\delta} + (\hat{\delta}_1^2 - 1).IV(f)_{11} + \textstyle\sum_{k=2}^{K}\delta_k^2 IV(f)_{kk} \end{aligned}$$

Since $C_{NT} \longrightarrow IV(f)$ as $p \longrightarrow \infty$ and since $\hat{\delta}$ is bounded ($\hat{\delta}'\hat{\delta} \leq 1$), $\hat{\delta}'\left(C_{NT} - IV(f)\right)\hat{\delta}$ is $o_p(1)$. Because $R^*(\underline{W}_1) - IV(f)_{11} \longrightarrow 0$ as $p \longrightarrow \infty$ (this result comes from the *step 10*, by taking $\hat{b}_1 = \underline{W}_1$) and $\hat{\delta}'\left(C_{NT} - IV(f)\right)\hat{\delta} \xrightarrow{p} 0$, we deduce that

$$(\hat{\delta}_1^2 - 1).IV(f)_{11} + \textstyle\sum_{k=2}^{K}\hat{\delta}_k^2.IV(f)_{kk} \xrightarrow{p} 0$$

The previous convergence result is obtained whatever $IV(f)$ is. Because $\forall k = 1, .., K$ $IV(f)_{kk} > 0$, we conclude that $\hat{\delta}_1^2 \longrightarrow 1$ and $\hat{\delta}_k^2 \longrightarrow 0$ $\forall k = 2, .., K$. Hence

$$\begin{aligned} S_1\frac{\underline{W'}_1\underline{b}_1}{p} &= \left|\frac{\underline{W'}_1\underline{b}_1}{p}\right| \\ &= \left|\left[b\left(\frac{b'b}{p}\right)^{-1/2}\hat{\delta} + \hat{V}\right]'\frac{\underline{b}_1}{p}\right| \\ &= \left|\left[\hat{\delta}'\left(\frac{b'b}{p}\right)^{-1/2}b' + \hat{V}'\right]\frac{\underline{b}_1}{p}\right| \\ &= \left|\hat{\delta}'\left(\frac{b'b}{p}\right)^{-1/2}\left(\frac{b'\underline{b}_1}{p}\right) + \hat{V}'\frac{\underline{b}_1}{p}\right| \end{aligned}$$

Since $\hat{V}'b = 0$, $\frac{b'\underline{b}_1}{p} \longrightarrow (1, 0, ..., 0)'$ and $\frac{b'b}{p} \longrightarrow I_K$ as $p \longrightarrow \infty$,

$$Plim \;\; S_1\frac{\underline{W'}_1\underline{b}_1}{p} = Plim \;\; \hat{\delta}_1$$

Because $\left(\hat{\delta}_1^2, ..., \hat{\delta}_K^2\right) \xrightarrow{p} (1, 0, ..., 0)$, it follows that $S_1\frac{\underline{W}_1'\underline{b}_1}{p} \longrightarrow 1$.
We use the same tricks to prove that for $k \in \{2, ..., K\}$, $Plim \;\; S_1\frac{\underline{W}_1'\underline{b}_k}{p} = 0$.
We conclude that $S_1\frac{\underline{W}_1' b}{p} \longrightarrow (1, 0, ..., 0) \equiv l_1'$

Step 12*:* We assume that the columns of W are formed by the K ordered eigenvectors of Σ, and is normalized as $\frac{W'W}{p} = I_K$. We define the matrix $S = Diag\left[sign(\underline{W}'_1\underline{b}_1), ..., sign(\underline{W}'_K\underline{b}_K)\right]$, where $\underline{A}_k$ is the k^{th} column of the matrix A. Then $S\frac{W'b}{p} \xrightarrow{p} I_K$.
To prove this result, we need to prove that for each column $\underline{W}_k$ of W, $S\frac{\underline{W}'_k b}{p} \longrightarrow (0, ..., 1, 0, ..., 0)$, with 1 corresponding to the position k. The result for the case of $k = 1$ is given by the *step 11*. The results for $k = 2, ..., K$ are based on steps 8 to 11, and consist on maximizing $R(.)$ and $R^*(.)$ in a sequential way, using orthonormal subspaces of Γ. For example, for the column $\underline{W}_k$ of W, we can write $\underline{W}_k = b\left(\frac{b'b}{p}\right)^{-1/2}\hat{\delta}_k + \hat{V}_k$, with $\hat{V}'_k b = 0$ and $\frac{\hat{V}'_k\hat{V}_k}{p} \xrightarrow{p} 0$ and $\hat{\delta}^2_{kl} \xrightarrow{p} 0$, $\forall\ \ l \neq k$ and $\hat{\delta}^2_{kk} \xrightarrow{p} 1$.

We proved through *steps 1* to *12* that $S\frac{W'b}{p} \xrightarrow{p} I_K$. This leads to $S\frac{W'b}{p}f_t \xrightarrow{p} f_t$. This result corresponds to the case where Σ is known. If Σ is unknown and is consistently estimated by $\hat{\Sigma}$, and if $\hat{W}$ is the matrix of ordered eigenvectors of $\hat{\Sigma}$ ($\hat{W}$ consistently estimates of W), then, we deduce that $S\frac{\hat{W}'b}{p}f_t \xrightarrow{p} f_t$.

B) Proof of: $\frac{1}{p}S\hat{W}'\epsilon_t \xrightarrow{p} 0$

S is defined as in the previous subsection. Note that for $k \in \{1, ..., K\}$

$$\begin{aligned}\frac{1}{p}S_k\hat{W}'_k\epsilon_t &= \frac{1}{p}\sum_{i=1}^{p} S_k\hat{W}'_{ik}\epsilon_{it}\\ &= \frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)\epsilon_{it} + \frac{1}{p}\sum_{i=1}^{p} b_{ik}\epsilon_{it}\end{aligned}$$

We are going to prove that $\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)\epsilon_{it} \xrightarrow{p} 0$ and $\frac{1}{p}\sum_{i=1}^{p} b_{ik}\epsilon_{it} \xrightarrow{p} 0$. By the Holder inequality

$$\left|\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)\epsilon_{it}\right| \leq \left[\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)^2\right]^{1/2} . \left[\frac{1}{p}\sum_{i=1}^{p}\epsilon^2_{it}\right]^{1/2}$$

In addition,

$$\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)^2 = S^2_k\frac{1}{p}\underline{\hat{W}}'_k\underline{\hat{W}}_k + \frac{1}{p}\sum_{i=1}^{p} b^2_{ik} - 2.\frac{1}{p}S_k\underline{\hat{W}}'_k\underline{b}_k$$

The convergence in probability to 0 of $\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)^2$ is deduced because

$$S^2_k\frac{1}{p}\underline{\hat{W}}'_k\underline{\hat{W}}_k \xrightarrow{p} 1,\ \frac{1}{p}\sum_{i=1}^{p} b^2_{ik} \xrightarrow{p} 1 \text{ and } \frac{1}{p}S_k\underline{\hat{W}}'_k\underline{b}_k \xrightarrow{p} 1.$$

Since $\frac{1}{p}\sum_{i=1}^{p}\epsilon^2_{it} \sim O_p(1)$, it follows that $\frac{1}{p}\sum_{i=1}^{p}\left(S_k\hat{W}_{ik} - b_{ik}\right)\epsilon_{it} \xrightarrow{p} 0$.
In other hand,

$$\begin{aligned}E\left[\left(\frac{1}{p}\sum_{i=1}^{p} b_{ik}\epsilon_{it}\right)^2\right] &= \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p} b_{ik}b_{jk}E(\epsilon_{it}\epsilon_{jt})\\ &\leq \frac{B^2}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p} E(\epsilon_{it}\epsilon_{jt}) \longrightarrow 0\end{aligned}$$

with B the bound of loadings. The last convergence result is justified by the fact that the loadings and $\frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|E(\epsilon_{it}\epsilon_{jt})|$ are bounded. Since the mean-squared convergence implies the convergence in probability, we conclude that $\frac{1}{p}\sum_{i=1}^{p} b_{ik}\epsilon_{it} \xrightarrow{p} 0$.

$\frac{1}{p}S\hat{W}'(u_t - u_{t-\Delta}) \xrightarrow{p} 0$

Let's assume that $(u_t - u_{t-\Delta})$ satisfies $A1$) and $A2$), then, using the same tricks as in the previous subsection, we can establish that $\frac{1}{p}S\hat{W}'(u_t - u_{t-\Delta}) \xrightarrow{p} 0$.

Proof of Lemma 4.1

Our estimator of the rotated factor is defined by

$$\hat{f}_{kt} = \frac{1}{p}W_k' r_t^* + \frac{1}{p}W_k'(u_t - u_{t-\Delta}) + \frac{1}{p}W_k^{\epsilon'} r_t^* + \frac{1}{p}W_k^{\epsilon'}(u_t - u_{t-\Delta})$$

Assume that p and n are sufficiently large such that $\frac{1}{p}W_k^{\epsilon'} r_t^*$ and $\frac{1}{p}W_k^{\epsilon'}(u_t - u_{t-\Delta})$ can be neglected. Then

$$\begin{aligned}
\hat{f}_{kt} &= \frac{1}{p}W_k' r_t^* + \frac{1}{p}W_k'(u_t - u_{t-\Delta}) + \frac{1}{p}W_k^{\epsilon'} r_t^* + \frac{1}{p}W_k^{\epsilon'}(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p}W_k' r_t^* + \frac{1}{p}W_k'(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p}W_k' b f_t + \frac{1}{p}W_k'\epsilon_t + \frac{1}{p}W_k'(u_t - u_{t-\Delta}) \\
&\approx \frac{1}{p}W_k' b f_t + \frac{1}{p}W_k'(u_t - u_{t-\Delta}) \\
&\approx \tilde{f}_{kt} + \frac{1}{p}W_k'(u_t - u_{t-\Delta})
\end{aligned}$$

The last two approximations are obtained since $\frac{1}{p}W_k'\epsilon_t = O_p(n^{-1/2}p^{-1/2})$ and p and n are assume to be large.

Since $\hat{f}_{kt} \approx \tilde{f}_{kt} + \frac{1}{p}W_k'(u_t - u_{t-\Delta})$, $E\left[\frac{1}{p}W_k'(u_t - u_{t-\Delta})|W_k', b, f\right] = 0$ and $\frac{1}{p}W_k'(u_t - u_{t-\Delta}) \perp \frac{1}{p}W_k'(u_s - u_{s-\Delta})$ $\forall s \neq t$, we deduce from the properties of the pre-averaging estimator of the integrated volatility (Jacod et Al.(2009)) that $PRV(\hat{f}_{kt})$ is an estimator of the integrated volatility of $\tilde{f}_{kt}$ with the rate of convergence of $n^{-1/4}$. We deduce that $\left[\tilde{f}_{kt}\right]^{\epsilon} \equiv PRV(\hat{f}_{kt}) - \left[\tilde{f}_{kt}\right] = O_p(n^{-1/4})$.

Proof of Lemma 4.2

Let $k \in 1,...,K$ and $i \in 1,...,p$. The estimator of the loading of asset i on factor k is defined by $\hat{b}_{ik} = \frac{MRC(r_i,\hat{f}_k)}{PRV(\hat{f}_k)}$. We are going firstly to establish the convergence rate of $\hat{b}_{ik} - b_{ik}$. Let's consider the two following notations:

$$MRC(r_i, \hat{f}_k) = \left[r_i^*, \tilde{f}_k\right] + \left[r_i^*, \tilde{f}_k\right]^{\epsilon}$$
$$PRV(\hat{f}_k) = \left[\tilde{f}_k\right] + \left[\tilde{f}_k\right]^{\epsilon}$$

where $[X]$ is the covariation of the process X, θ^{ϵ} is the estimation error in the estimation of θ. Using these notations, we obtain

$$\begin{aligned}\hat{b}_{ik} &= \frac{\left[r_i^*,\tilde{f}_k\right]+\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]+\left[\tilde{f}_k\right]^{\epsilon}} \\ &= \left(\frac{\left[r_i^*,\tilde{f}_k\right]+\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)\left(1+\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^{-1} \\ &= \left(\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}+\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)\left(1+\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^{-1}\end{aligned}$$

Since $\left[\tilde{f}_k\right]^{\epsilon}$ is the error in the estimation of $\left[\tilde{f}_k\right]$ using the pre-averaging estimator $PRV(\hat{f}_k)$, we can assume that $\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}$ is closed to 0, such that the following Taylor expansion holds

$$\left(1+\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^{-1} = 1-\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}+O_p\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)$$

Then

$$\begin{aligned}\hat{b}_{ik} &= \left(\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}+\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)\left(1-\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}+O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)\right) \\ &= \frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}-\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}\cdot\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}+\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right) \\ &\quad +\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}-\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\cdot\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]^2}+\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)\end{aligned}$$

It follows that

$$\hat{b}_{ik}-\tilde{b}_{ik} = -\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}\cdot\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}+\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)+\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}-\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\cdot\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]^2}+\frac{\left[r_i^*,\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)$$

Then

$$\begin{aligned}\left|\hat{b}_{ik}-\tilde{b}_{ik}\right| &\leq \left|\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]^2}\right|\cdot\left|\left[\tilde{f}_k\right]^{\epsilon}\right|+\left|\frac{\left[r_i^*,\tilde{f}_k\right]}{\left[\tilde{f}_k\right]}\right|\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right) \\ &\quad +\frac{1}{\left[\tilde{f}_k\right]}\cdot\left|\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\right|+\frac{1}{\left[\tilde{f}_k\right]^2}\cdot\left|\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\right|\cdot\left|\left[\tilde{f}_k\right]^{\epsilon}\right|+\frac{1}{\left[\tilde{f}_k\right]}\cdot\left|\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\right|\cdot O\left(\left(\frac{\left[\tilde{f}_k\right]^{\epsilon}}{\left[\tilde{f}_k\right]}\right)^2\right)\end{aligned}$$

By the $Lemma 4.1$ $\left[\tilde{f}_k\right]^{\epsilon} = O_p(n^{-1/4})$. In addition, $r_i = r_i^* + (u_t - u_{t-\Delta})$ and $\hat{f}_{kt} \approx \tilde{f}_{kt} + \frac{1}{p}W_k'(u_t - u_{t-\Delta})$. It follows from the pre-averaging estimator of the integrated covariation of Kim Christensen et Al.(2010) that $\left|\left[r_i^*,\tilde{f}_k\right]^{\epsilon}\right| \equiv \left|MRC(r_i,\hat{f}_k)\right| = O_p(n^{-1/4})$. Hence

$$\begin{aligned}\left|\hat{b}_{ik}-\tilde{b}_{ik}\right| &\leq O_p(n^{-1/4})+O_p(1)O(O_p(1)O_p(n^{-1/4})^2)+O_p(1)O_p(n^{-1/4}) \\ &\quad +O_p(1)O_p(n^{-1/4})O_p(n^{-1/4})+O_p(1)O_p(n^{-1/4})O(O_p(1)O_p(n^{-1/4})^2) \\ &\leq O_p(n^{-1/4})+O_p(n^{-1/2})+O_p(n^{-1/4})+O_p(n^{-1/2})+O_p(n^{-1/4})O_p(n^{-1/2}) \\ &\leq O_p(n^{-1/4})\end{aligned}$$

Hence $\left|\hat{b}_{ik}-\tilde{b}_{ik}\right| = O_p\left(n^{-1/4}\right)$, $\forall k = 1, ..., K$, $\forall i = 1, ..., p$.
Using the Frobenius norm, we obtain

$$\begin{aligned}\left\|\hat{b}_k-\tilde{b}_k\right\|_F^2 &= \textstyle\sum_{i=1}^p \left|\hat{b}_{ik}-\tilde{b}_{ik}\right|^2 \\ &= \textstyle\sum_{i=1}^p O_p(n^{-1/2}) \\ &= O_p(pn^{-1/2})\end{aligned}$$

We conclude that $\left\|\hat{b}_k-\tilde{b}_k\right\|_F = \left\|\hat{b}_k-\tilde{b}_k\right\|_2 = O_p(p^{1/2}n^{-1/4})$

Proof of Lemma 4.3

We define the estimator of the integrated volatility of the idiosyncratic error terms by

$$\forall i = 1, ..., p,\ \hat{\Sigma}^{\epsilon}_{ii} = PRV(\hat{\epsilon}_i)$$

with $\hat{\epsilon}_{it} = r_{it} - \hat{b}_i \hat{f}_t$.
It can be easily established that

$$\hat{\epsilon}_{it} = \epsilon_{it} + (u_t - u_{t-\Delta}) - \sum_{k=1}^{K} \tilde{b}_{ik}\tilde{f}^{\epsilon}_{kt} - \sum_{k=1}^{K} \tilde{b}^{\epsilon}_{ik}\tilde{f}_{kt} - \sum_{k=1}^{K} \tilde{b}^{\epsilon}_{ik}\tilde{f}^{\epsilon}_{kt}$$

Since $\hat{f}_{kt}$ is a consistent estimator of $\tilde{f}_{kt}$, let's assume that n and p are sufficiently large such that $\tilde{f}^{\epsilon}_{kt} \equiv \hat{f}_{kt} - \tilde{f}_{kt}$ can be neglected. Then

$$\hat{\epsilon}_{it} = \epsilon_{it} + (u_t - u_{t-\Delta}) + O_p(n^{-3/4})$$

It follows that

$$PRV(\hat{\epsilon}_i) = [\epsilon_i] + O_p(n^{-1/4})$$

Hence

$$\left|\hat{\Sigma}^{\epsilon}_{ii} - \Sigma^{\epsilon}_{ii}\right| = O_p(n^{-1/4})$$

Under the Frobenius norm

$$
\begin{aligned}
\left\|\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\|_F^2 &= \sum_{i=1}^{p}\left|\hat{\Sigma}^{\epsilon}_{ii}-\Sigma^{\epsilon}_{ii}\right|^2 \\
&= \sum_{i=1}^{p} O_p(n^{-1/2}) \\
&= O_p(pn^{-1/2})
\end{aligned}
$$

We conclude that $\left\|\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\|_F = O_p(p^{1/2}n^{-1/4})$.

Proof of Theorem 4.1

$$
\begin{aligned}
\left\|\hat{\Sigma}-\Sigma\right\| &= \left\|\sum_{k=1}^{K}\left(\hat{b}_k\hat{b}_k'\hat{\Sigma}^f_{kk}-b_kb_k'\Sigma^f_{kk}\right)+\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\| \\
&\leq \left\|\sum_{k=1}^{K}\left(\hat{b}_k\hat{b}_k'\hat{\Sigma}^f_{kk}-b_kb_k'\Sigma^f_{kk}\right)\right\|+\left\|\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\| \\
&\leq \sum_{k=1}^{K}\left[\left\|\left(\hat{b}_k-b_k\right)\left(\hat{b}_k-b_k\right)'\right\|+\left\|\left(\hat{b}_k-b_k\right)b_k'\right\|+\left\|b_k\left(\hat{b}_k-b_k\right)'\right\|\right]\cdot\left|\hat{\Sigma}^f_{kk}-\Sigma^f_{kk}\right| \\
&\quad +\left[\left\|\left(\hat{b}_k-b_k\right)\left(\hat{b}_k-b_k\right)'\right\|+\left\|\left(\hat{b}_k-b_k\right)b_k'\right\|+\left\|b_k\left(\hat{b}_k-b_k\right)'\right\|\right]\cdot\Sigma^f_{kk}+\left\|\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\| \\
&\leq \sum_{k=1}^{K}\left\|\hat{b}_k-b_k\right\|^2\cdot\left|\hat{\Sigma}^f_{kk}-\Sigma^f_{kk}\right|+2\sum_{k=1}^{K}\left\|\hat{b}_k-b_k\right\|\cdot\|b_k'\|\cdot\left|\hat{\Sigma}^f_{kk}-\Sigma^f_{kk}\right| \\
&\quad +\sum_{k=1}^{K}\|b_k\|^2\cdot\left|\hat{\Sigma}^f_{kk}-\Sigma^f_{kk}\right|+\sum_{k=1}^{K}\left\|\hat{b}_k-b_k\right\|^2\Sigma^f_{kk} \\
&\quad +2\sum_{k=1}^{K}\left\|\hat{b}_k-b_k\right\|\cdot\|b_k\|\,\Sigma^f_{kk}+\left\|\hat{\Sigma}^{\epsilon}-\Sigma^{\epsilon}\right\| \\
&\leq \sum_{k=1}^{K}O_p(pn^{-1/4})O_p(n^{-1/4})+\sum_{k=1}^{K}O_p(p^{1/2}n^{-1/4})O_p(p^{1/2}) \\
&\quad +\sum_{k=1}^{K}O_p(p^{1/2})O_p(p^{1/2}n^{-1/4})O_p(n^{-1/4})+\sum_{k=1}^{K}O_p(p)O_p(n^{-1/4}) \\
&\quad +\sum_{k=1}^{K}O_p(pn^{-1/2})O_p(1)+\sum_{k=1}^{K}O_p(p^{1/2}n^{-1/4})O_p(p^{1/2})O_p(1) \\
&\quad +\sum_{k=1}^{K}O_p(p^{1/2})O_p(p^{1/2}n^{-1/4})O_p(1)+O_p(p^{1/2}n^{-1/4}) \\
&\leq O_p(Kpn^{-1/4})
\end{aligned}
$$

Concerning the multidimensional pre-averaging estimator MRC^{δ} and the multidimensional kernel estimator $MRker$, we use the rate of convergence of each of their elements ($n^{-1/5}$, from Christensen et Al. (2010) and Barndorff-Nielsen et Al. (2008)) to obtain that:
a) $\left\|MRC^{\delta}-\Sigma\right\| = O_p\left(pn^{-1/5}\right)$;
b) $\|MRker-\Sigma\| = O_p\left(pn^{-1/5}\right)$.
The steps of the proof are the same as for the factorial case, the only difference is that

$\forall i, j = 1, ..., p$, $\left| MRC_{ij}^{\delta} - \Sigma_{ij} \right| = O_p\left(n^{-1/5}\right)$ and $\forall i, j = 1, ..., p$, $\left| MRKer_{ij} - \Sigma_{ij} \right| = O_p\left(n^{-1/5}\right)$.