Biometrika (2017), **104**, 2, *pp.* 397–410
Printed in Great Britain
doi: 10.1093/biomet/asx014
Advance Access publication 21 April 2017

On the number of common factors with high-frequency data

By XIN-BING KONG

School of Science, Nanjing Audit University, West Yushan Road, Nanjing 211815, China
kongxblqh@gmail.com

Summary

In this paper, we introduce a local principal component analysis approach to determining the number of common factors of a continuous-time factor model with time-varying factor loadings using high-frequency data. The model is approximated locally on shrinking blocks using discrete-time factor models. The number of common factors is estimated by minimizing the penalized aggregated mean squared residual error over all shrinking blocks. While the local mean squared residual error on each block converges at rate $\min(n^{1/4}, p)$, where n is the sample size and p is the dimension, the aggregated mean squared residual error converges at rate $\min(n^{1/2}, p)$; this achieves the convergence rate of the penalized criterion function of the global principal component analysis method, assuming restrictive constant factor loading. An estimator of the number of factors based on the local principal component analysis is consistent. Simulation results justify the performance of our estimator. A real financial dataset is analysed.

Some key words: Continuous-time factor model; High-frequency data; Itô process.

1. Introduction

Itô processes are widely used in biology (Fleming, 1975; Erban & Othmer, 2014) and finance (Barndorff-Nielsen & Shephard, 2002). Recent years have seen rapid progress in inference for the characteristics of finite-dimensional Itô processes. Andersen et al. (2003) and Barndorff-Nielsen & Shephard (2002) analysed the realized variance of stochastic volatility models. Mykland & Zhang (2006) investigated regression relations between two Itô processes. Li & Mykland (2008) studied the robustness of the realized variance in estimating the integrated volatility. Zhang et al. (2005) and Jacod et al. (2009) developed two-time-scale and pre-averaging techniques to estimate the integrated volatility of a univariate Itô process using noisy data. Pre-averaging was later extended by Christensen et al. (2010) to estimate integrated volatility matrices of finite dimension. Todorov & Tauchen (2012a,b) investigated the distribution of the volatility using the realized Laplace transform. For related work, see Mykland & Zhang (2009).

Owing to the rapid development of data collection technology, we now have to deal with vast real-time data streams. Large-dimensional Itô processes have become a natural tool for modelling such data with serial and cross-sectional dependence. Let $X_t = (X_{1t}, \ldots, X_{pt})^{\mathrm{T}}$, where X_{it} is the log-price process of the ith asset defined on the filtered probability space $\{\Omega, \mathcal{F}, \mathcal{F}_t, P\}$. Although there are a large number of assets, our model assumes that they are driven by common factors $Y_t = (Y_{1t}, \ldots, Y_{r_0 t})^{\mathrm{T}}$ and other specific factors Z_{it}; that is,

$$\mathrm{d}X_t = \Lambda_t \, \mathrm{d}Y_t + \mathrm{d}Z_t, \tag{1}$$

where Λ_t is a $p \times r_0$ matrix of continuous-time factor loading processes. Model (1) in the high-dimensional setting was first studied under the assumption that $\Lambda_t = \Lambda$ is a constant matrix by Aït-Sahalia & Xiu (2017) and in an unpublished 2016 Stanford University technical report by M. Pelger. Interesting factor analysis results were reported in these papers. However, the constancy assumption on Λ_t restricts the flexibility of the systematic volatilities. For example, if $r_0 = 1$, the component volatility processes of $\Lambda\,\mathrm{d}Y_t$ are perfectly correlated. In contrast, the general model (1) allows independence between rows of the volatility matrix of $\Lambda_t\,\mathrm{d}Y_t$. In this paper, as in Aït-Sahalia & Xiu (2017), we assume that Z_t and Y_t are continuous semi-martingales driven by Brownian motions. The technical report of Pelger allows for the presence of jump components in Y_t and Z_t. Under model (1), adding jumps to both the driving processes (Y_t, Z_t) and the high-dimensional matrix of factor loading processes Λ_t significantly complicates the mathematical technicalities. We leave the treatment of jumps to future work.

For ease of identifying the systematic and idiosyncratic volatility processes, we write (1) equivalently as

$$X_{it} = X_{i0} + \int_0^t \mu_{is}\,\mathrm{d}s + \int_0^t \sigma_{is}^1\,\mathrm{d}W_s^1 + \cdots + \int_0^t \sigma_{is}^{r_0}\,\mathrm{d}W_s^{r_0} + \int_0^t \sigma_{is}^*\,\mathrm{d}W_{is}^* \quad (i = 1, \ldots, p), \tag{2}$$

where μ_{is}, σ_{is}^l $(l = 1, \ldots, r_0)$ and σ_{is}^* are locally bounded adapted processes, and W_s^l $(1 \leqslant l \leqslant r_0)$ and W_{is}^* are standard Brownian motions. In matrix form, (2) is

$$\mathrm{d}X_t = B_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t + \sigma_t^*\,\mathrm{d}W_t^*,$$

where $B_t = (\mu_{1t}, \ldots, \mu_{pt})^{\mathrm{T}}$, $\sigma_t^* = \mathrm{diag}(\sigma_{1t}^*, \ldots, \sigma_{pt}^*)$, $W_t = (W_t^1, \ldots, W_t^{r_0})^{\mathrm{T}}$, $W_t^* = (W_{1t}^*, \ldots, W_{pt}^*)^{\mathrm{T}}$ and $\sigma_t = (\sigma_t^1, \ldots, \sigma_t^{r_0})$ with $\sigma_t^l = (\sigma_{1t}^l, \ldots, \sigma_{pt}^l)^{\mathrm{T}}$. For identification purposes, we assume throughout the paper that $\mathrm{cov}(W_1) = I_{r_0}$. We also assume that W_t and W_t^* are independent, though this condition can be relaxed to weak dependence. Barndorff-Nielsen et al. (2008) and Fan et al. (2012) used (2) to mimic the real log-prices in their simulation studies.

In this paper, we aim to estimate r_0 under model (2) using high-frequency data when both p and n are large. The availability of high-frequency data facilitates statistical inference of volatilities, as stationarity of the data-generating process and the volatility process is not required. In Aït-Sahalia & Xiu (2017) and the 2016 technical report of M. Pelger, consistent estimators of r_0 were proposed for the case where $\Lambda_t = \Lambda$. Aït-Sahalia & Xiu (2017) estimates r_0 via minimizing penalized eigenvalues of the integrated volatility matrix, while the report of M. Pelger does so by ranking the perturbed ratios of two adjacent eigenvalues. Although they use different ranking objectives, both papers are based on global principal component analysis of the integrated volatility matrix. A prerequisite for global principal component analysis is constancy of the factor loading matrix.

In contrast, in the present paper, we aggregate the local factor analysis results to give an alternative estimator that is robust with respect to the time-varying factor loading matrix. We approximate (2) locally using discrete factor models introduced in Chamberlain & Rothschild (1983) on non-overlapping shrinking time windows. We then calculate the local mean squared residual errors assuming r common driving processes. An estimator of r_0 is obtained by minimizing the aggregated mean squared residual error with respect to r. The key to success in obtaining a consistent estimator is that σ_t is approximately constant over a shrinking local window. The aggregation of local analysis of high-frequency data with a time-varying volatility process arises frequently in estimation of low-dimensional volatility functionals (Jacod & Rosenbaum, 2013; Li et al., 2013). However, aggregation of high-dimensional local factor analysis is still not well

understood. Indeed, our theory demonstrates that while the local mean squared residual error converges at rate $\min(p, n^{1/4})$, the aggregated mean squared residual error converges at a rate of $\min(p, n^{1/2})$.

Ignoring the discretization error, local factor analysis within shrinking blocks can be done using the standard method for discrete-time factor models. There are useful results in the literature on high-dimensional discrete-time factor models with a constant and deterministic loading matrix. Bai & Ng (2002) developed two criteria for determining the number of common factors; Stock & Watson (2002) investigated the problem of forecasting using estimated principal components of an approximate factor model; Onatski (2012) proposed an alternative estimator using differenced eigenvalues; and Ahn & Hohenstein (2013) derived two estimators of the number of factors by maximizing the ratio of two adjacent eigenvalues. Our local analysis method is borrowed from Bai & Ng (2002) and evaluates the mean squared residual error; this is because local mean squared residual errors are easy to aggregate, whereas the differences (Onatski, 2012) or ratios (Ahn & Hohenstein, 2013) between eigenvalues of local volatility matrices are inconvenient, if not impossible, to obtain.

2. Methodology

We assume that the dataset is discretely sampled from the high-dimensional Itô process X_t with equal sampling length $\Delta_n = T/n$, where T is the time horizon and n is the sample size. Mathematically, we consider the asymptotic regime where $\Delta_n \to 0$ with T fixed. Let X_{it_k} be the observation of X_{it} at time t_k, with $\Delta_n = t_k - t_{k-1}$ and $\Delta_j^n X_i = X_{it_j} - X_{it_{j-1}}$ being the one-step increment. We separate the sampling time-points $\{0 = t_0 < t_1 < \cdots < t_n = T\}$ into $\lfloor n/k_n \rfloor$ non-overlapping blocks, with each block containing $k_n = \lfloor \theta n^{1/2} \rfloor$ one-step increments, where $\lfloor x \rfloor$ stands for the largest integer smaller than or equal to x and θ is a constant. This choice of k_n balances the discretization bias, approximation accuracy and aggregation efficiency under the smoothness conditions presented in § 3. We discuss the choice of θ for finite n in § 4.

Restricted to the kth block, for $j = 1, \ldots, k_n$, $i = 1, \ldots, p$ and $l = 1, \ldots, r_0$ we let $\delta_k = \Delta_n^{-1/2}(\Delta^n_{(k-1)k_n+j} X_i) = (\delta^k_{ji})_{k_n \times p}$ be a $k_n \times p$ matrix, $\mu_k = (\mu_{it_{(k-1)k_n+j}})$ a $p \times k_n$ matrix, $F_k^0 = \Delta_n^{-1/2}(\Delta^n_{(k-1)k_n+j} W^l)$ an $r_0 \times k_n$ matrix, $\sigma_k^0 = (\sigma^l_{it_{(k-1)k_n}})$ a $p \times r_0$ matrix, $\sigma_k^* = \mathrm{diag}\{\sigma^*_{1t_{(k-1)k_n}}, \ldots, \sigma^*_{pt_{(k-1)k_n}}\}$ a $p \times p$ diagonal matrix, and $F_k^* = \Delta_n^{-1/2}(\Delta^n_{(k-1)k_n+j} W_i^*) = (F_k^*(1), \ldots, F_k^*(k_n))$ a $p \times k_n$ matrix. Then, as $k_n \Delta_n$ shrinks to zero, we expect that

$$\delta_k^{\mathrm{T}} \doteq \mu_k \Delta_n^{1/2} + \sigma_k^0 F_k^0 + \sigma_k^* F_k^* = \mu_k \Delta_n^{1/2} + \bar{\delta}_k^{\mathrm{T}}, \tag{3}$$

where $\doteq$ means approximate equality. The discretization error in (3) will be shown to be negligible in determining r_0 under the regularity conditions given in § 3. Now the right-hand side of (3) has the exact structure of a discrete approximate factor model, where $\mu_k \Delta_n^{1/2}$ is a negligible mean, σ_k^0 is the factor loading matrix fixed in block k but varying across k, the F_k^0 are the common factors, and F_k^* and σ_k^* are the specific factors and their loadings, respectively.

We determine r_0 by trial and error. Suppose that the number of common driving processes we try is $r < r_m$ for some predetermined finite number r_m. Let $\sigma_k^r = (\lambda^r_{k1}, \ldots, \lambda^r_{kp})^{\mathrm{T}}$ and $F_k^r = \{F_k^r(1), \ldots, F_k^r(k_n)\}$ be the matrices consisting of the first r columns of σ_k^0 and the first r rows of F_k^0, respectively. Then, as in Bai & Ng (2002), we estimate σ_k^r and F_k^r by solving the

optimization problem

$$V(r) = \min_{\sigma_k^r, F_k^r} (pk_n)^{-1} \sum_{i=1}^{p} \sum_{j=1}^{k_n} \{\delta_{ji}^k - (\lambda_{ki}^r)^{\mathrm{T}} F_k^r(j)\}^2, \tag{4}$$

subject to $F_k^r(F_k^r)^{\mathrm{T}}/k_n = I_r$ or $(\sigma_k^r)^{\mathrm{T}}\sigma_k^r/p = I_r$, which are identifiability conditions commonly used in factor analysis. If the condition $F_k^r(F_k^r)^{\mathrm{T}}/k_n = I_r$ is used, the solution to (4) is $(\delta_k^{\mathrm{T}}\bar{F}_k^r/k_n, \bar{F}_k^r)$, where $(\bar{F}_k^r)_{r\times k_n}$ is the transpose of $\sqrt{k_n}$ times the eigenvector matrix of $\delta_k\delta_k^{\mathrm{T}}/(pk_n)$ corresponding to the r largest eigenvalues. If the second restriction is imposed, the solution to (4) is $\{\bar{\sigma}_k^r, (\bar{\sigma}_k^r)^{\mathrm{T}}\delta_k^{\mathrm{T}}/p\}$, where $(\bar{\sigma}_k^r)_{p\times r}$ is $\sqrt{p}$ times the eigenvector matrix of $\delta_k^{\mathrm{T}}\delta_k/(pk_n)$ corresponding to the r largest eigenvalues. The choice of condition depends on the magnitudes of k_n and p. If $k_n < p$, the former is used; otherwise the latter is used for less costly computation. Supposing that the first restriction is used, as in Bai & Ng (2002), we estimate F_k^0 by $\hat{F}_k^r = \bar{F}_k^r\delta_k\delta_k^{\mathrm{T}}/(pk_n)$. Then, the mean squared residual risk when r is tried and $\hat{F}_k^r$ serves as the design matrix in the regression $\delta_k^{\mathrm{T}} = \sigma_k^r\hat{F}_k^r + \tilde{\epsilon}_k$ equals

$$V(r, \hat{F}_k^r) = (pk_n)^{-1} \operatorname{tr}\big(\delta_k^{\mathrm{T}}[I_{k_n} - (\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r]\delta_k\big).$$

Letting $\hat{F}^r = (\hat{F}_1^r, \ldots, \hat{F}_{\lfloor n/k_n\rfloor}^r)^{\mathrm{T}}$, we define the aggregated mean squared residual risk as

$$U(r, \hat{F}^r) = \lfloor n/k_n\rfloor^{-1} \sum_{k=1}^{\lfloor n/k_n\rfloor} V(r, \hat{F}_k^r).$$

Ignoring the discretization error in (3), the theory in Bai & Ng (2002) implies that $V(r, \hat{F}^r)$ converges at a rate of $\min(k_n^{1/2}, p)$. But our theory demonstrates that $U(r, \hat{F}_k^r)$ converges at the higher rate of $\min(n^{1/2}, p)$. This indicates that the aggregation improves our estimates of r_0. Our technical proof shows that after proper centring, $U(r, \hat{F}^r)$ can be approximated by a martingale.

Minimizing $U(r, \hat{F}^r)$ with respect to r alone results in a large estimated r_0 due to overfitting. We therefore estimate r_0 by minimizing the following criterion function with a penalty:

$$\hat{r}_0 = \arg\min_{r\leqslant r_m} U_p(r) = \arg\min_{r\leqslant r_m} \{U(r, \hat{F}^r) + \beta r g(p, n)\}, \tag{5}$$

where $g(p, n)$ is a function of p and n, and β is a tuning parameter. The global principal component analysis method corresponds to setting $k_n = n$ in (5). It works for model (1) with constant Λ_t but does not work for model (2), since the discretization error of the factor loading process stands out.

A rule of thumb for tuning β in (5) is as follows. As in Bai & Ng (2002), we set β to be an estimate of $\zeta(\theta) = \lfloor n/k_n\rfloor^{-1}p^{-1}\sum_{k=1}^{\lfloor n/k_n\rfloor}\sum_{i=1}^{p}\sigma_k^{*2}(i,i)$, the averaged volatility of the error process. Here the dependence on θ is via $k_n = \lfloor\theta\sqrt{n}\rfloor$. If there is prior knowledge of r_0, denoted by $\tilde{r}_0$, an estimate of $\zeta(\theta)$ is $\hat{\zeta}(\theta) = U(\tilde{r}_0, \hat{F}^{\tilde{r}_0})$. In financial econometrics, a typical prior estimate of r_0 is 3 or 4, by virtue of the Fama–French factor model (Fama & French, 1992). If there is no prior estimate, as suggested in Bai & Ng (2002) we set $\hat{\zeta}(\theta) = U(r_m, \hat{F}^{r_m})$.

In the present paper, for simplicity we exclude microstructure noise and asynchronicity; but to reduce possible biases from them, we use relatively sparse high-frequency data, such as five-minute or ten-minute returns, instead of tick-by-tick ultrahigh-frequency data. In the context of estimating volatilities, Aït-Sahalia et al. (2005) indicated that sampling lengths of up to five minutes are safe for avoiding the effects of microstructure noise. Similar empirical findings were reported in Fan et al. (2016). With relatively sparse sampling, the asynchronicity bias is significantly eliminated since the time lag between the observation time and the closest trading time is very small relative to the sampling length, especially when the assets are heavily traded. In our numerical studies, we use a sampling length of five or ten minutes. Finally, we remark that the estimator based on aggregating local factor analysis remains consistent even if the observations are unequally spaced, as long as all local windows shrink to zero. However, for neatness of presentation, empirical application and mathematical proofs, we stick to the equal sampling scheme in this paper.

3. Assumptions

To establish the theoretical results, we need the following technical assumptions. The first assumption gives some regularity conditions on the coefficient processes of the Itô processes. These are commonly used in the literature (see, e.g., Wang & Zou, 2010; Jing et al., 2012a,b; Jacod & Podolskij, 2013; Jacod & Rosenbaum, 2013).

Assumption 1. We have a sequence T_n of stopping times that increase to infinity and a sequence a_n of numbers such that when $t < T_n$, $|Z_t| \leqslant a_n$ for $Z = \mu_i, \sigma_i^l, \sigma_i^*$ $(i = 1, \ldots, p;\ l = 1, \ldots, r_0)$ and

$$\left|E_{\mathcal{F}_{\min(t,\tau_n)}}\{Z_{\min(t+s,\tau_n)} - Z_{\min(t,\tau_n)}\}\right| + E_{\mathcal{F}_{\min(t,\tau_n)}}\{Z_{\min(t+s,\tau_n)} - Z_{\min(t,\tau_n)}\}^2 \leqslant a_n s.$$

Assumption 1 is mainly used to control the discretization error in each shrinking block. The class of continuous Itô processes driven by Brownian motion with locally bounded drift and volatility processes satisfies this condition. The next assumption gives a condition on the cross-sectional dependence of the specific driving Brownian motions.

Assumption 2. We have that $\log p = o(k_n^{1-\epsilon})$ and $\sqrt{n}\, p^{-2\delta'} = o(1)$ for some $\delta' \geqslant 1$ and any $\epsilon > 0$, and that $p^{-1} \sum_{i=1}^p \sum_{j=1}^p |\rho_{ij}^*| < M$ for some $M < \infty$, where ρ_{ij}^* is the correlation of W_{it}^* and W_{jt}^*.

The intuition behind Assumption 2 is that the specific factors are cross-sectionally weakly dependent. It includes the strict factor model as a special case. It is also fulfilled when the components of W^* are m-dependent or satisfy some mixing conditions. Assumption 2 implies that $E\{p^{-1/2} \sum_{i=1}^p F_k^*(i,1)\}^{2\delta'} \leqslant C(p^{-1} \sum_i \sum_j |\rho_{ij}^*|)^{\delta'} < M$ and $E[p^{-1/2} \sum_{i=1}^p \{F_k^{*2}(i,1) - 1\}]^{2\delta'} \leqslant C\{p^{-1} \sum_i^p \sum_j^p (\rho_{ij}^*)^2\}^{\delta'} < M$ for all $k = 1, \ldots, \lfloor n/k_n \rfloor$.

The next assumption is on the minimum eigenvalue of $\sigma_t^{\mathrm{T}} \sigma_t / p$. It is a continuous-time analogue of Assumption B in Bai & Ng (2002) and Assumption 1 in Fan et al. (2013).

Assumption 3. There exists a sequence of constants $b_n > 0$ such that $\inf_{0 \leqslant t \leqslant T_n} \lambda_{\min}(p^{-1} \sigma_t^{\mathrm{T}} \sigma_t) \geqslant b_n$ and, for all $t \in [0, T]$, $p^{-1} \sigma_t^{\mathrm{T}} \sigma_t$ almost surely has distinct eigenvalues, where $\lambda_{\min}(A)$ stands for the minimum eigenvalue of A.

4. MAIN RESULTS

From (3), (A1), (A2) and the independence between F_k^0 and F_k^*, we have that $\delta_k \delta_k^{\mathrm{T}} \doteq (\sigma_k^0 F_k^0)^{\mathrm{T}} \sigma_k^0 F_k^0$. One therefore expects that $\hat{F}_k^r \doteq H_k^r F_k^0$, where $H_k^r = (pk_n)^{-1} \bar{F}_k^r (\sigma_k^0 F_k^0)^{\mathrm{T}} \sigma_k^0$. Ignoring the discretization error, the theory in Bai & Ng (2002) demonstrates that $V(r, \hat{F}_k^r) - V(r, H_k^r F_k^0) = O_{\mathrm{p}}[\{\min(\sqrt{p}, n^{1/4})\}^{-1}]$ for each k. However, the convergence rate of the aggregated error $\lfloor n/k_n \rfloor^{-1} \sum_{k=1}^{\lfloor n/k_n \rfloor} \{V(r, \hat{F}_k^r) - V(r, H_k^r F_k^0)\}$ is unclear from the literature. The difficulty is that $\bar{F}_k^r$ contained in H_k^r depends on $\mathcal{F}_{t_{kk_n}}$ and has no explicit dependence on $\mathcal{F}_{t_{(k-1)k_n}}$ and F_k^0. Nevertheless, our first result below shows that $\bar{F}_k^r$ approximately depends only on $\mathcal{F}_{t_{(k-1)k_n}}$ and F_k^0, which is crucial in deriving the convergence rate of the aggregated mean squared residual error.

THEOREM 1. *Suppose that $k_n = \lfloor \theta \sqrt{n} \rfloor$. Let L_k be the diagonal matrix consisting of the eigenvalues of $B_k = (\sigma_k^0)^{\mathrm{T}} \sigma_k^0 / p$, and let γ_k be the eigenvector matrix of B_k. Then, under Assumptions 1–3,*

$$E_{\mathcal{F}_{k-1}} \left\{ \| (pk_n^2)^{-1} \bar{F}_k^r (F_k^0)^{\mathrm{T}} (\sigma_k^0)^{\mathrm{T}} \sigma_k^0 F_k^0 (\bar{F}_k^r)^{\mathrm{T}} - L_k \| \right\} \leqslant C C_{pn}^{-1}, \tag{6}$$

$$\max_{1 \leqslant k \leqslant \lfloor n/k_n \rfloor} E_{\mathcal{F}_{k-1}} \left[\| F_k^0 (\bar{F}_k^r)^{\mathrm{T}} / k_n - \{(\sigma_k^0)^{\mathrm{T}} \sigma_k^0 / p\}^{-1/2} \gamma_k L_k^{1/2} \| \right] = O_{\mathrm{p}}(C_{pn}^{-1}), \tag{7}$$

$$\max_{1 \leqslant k \leqslant \lfloor n/k_n \rfloor} E_{\mathcal{F}_{k-1}} \left\{ \| \bar{F}_k^r / \sqrt{k_n} - \gamma_k^{\mathrm{T}} F_k^0 / \sqrt{k_n} \| \right\} = O_{\mathrm{p}} \left(C_{pn}^{-1} \right), \tag{8}$$

where $C_{pn} = \min(\sqrt{p}, n^{1/4})$, C is some constant, and $\|A\|$ stands for the spectral norm of a matrix A.

Remark 1. From (7), $F_k^0 (\bar{F}_k^r)^{\mathrm{T}}$ can be approximated by a measurable function on $\mathcal{F}_{t_{(k-1)k_n}}$; (8) implies that the $\bar{F}_k^r$ ($k = 1, \ldots, \lfloor n/k_n \rfloor$) approximately form a sequence of martingale differences with respect to $\mathcal{F}_{t_k}$ ($k = 1, \ldots, \lfloor n/k_n \rfloor$). We conjecture that, as functions of $\bar{F}_k^r$, the $\{V(r, \hat{F}_k^r) - V(r, H_k^r F_k^0)\}$ ($k = 1, \ldots, \lfloor n/k_n \rfloor$) approximately form a sequence of martingale differences as well.

The next theorem reveals that the aggregated mean squared residual error $U(r, \hat{F}^r)$ is asymptotically strictly greater than $U(r_0, F^0)$ when $r < r_0$, where $F^0 = (F_1^0, \ldots, F_{\lfloor n/k_n \rfloor}^0)^{\mathrm{T}}$, but is close to $U(r_0, F^0)$ when $r > r_0$. Let $H^r = (H_1^r, \ldots, H_{\lfloor n/k_n \rfloor}^r)^{\mathrm{T}}$ and define $H^r \circ F^0 = (H_1^r F_1^0, \ldots, H_{\lfloor n/k_n \rfloor}^r F_{\lfloor n/k_n \rfloor}^0)^{\mathrm{T}}$.

THEOREM 2. *Suppose that $k_n = \lfloor \theta \sqrt{n} \rfloor$. Under Assumptions 1–3, we have that for $r \leqslant r_0$, $U(r, \hat{F}^r) - U(r, H^r \circ F^0) = O_{\mathrm{p}}(p^{-1} \vee n^{-1/2})$; and for $r_m > r > r_0$, $U(r, \hat{F}^r) - U(r_0, F^0) = O_{\mathrm{p}}(p^{-1} \vee n^{-1/2})$. If $r < r_0$, then for some $\tau > 0$ we have $\liminf_{p,n \to \infty} U(r, H^r \circ F^0) - U(r_0, F^0) = \tau$ in probability.*

Remark 2. In block k, with the discretization error included, our proof implies that $V(r, \hat{F}_k^r) - V(r, H_k^r F_k^0) = O_{\mathrm{p}}\{\max(p^{-1}, n^{-1/4})\}$. However, Theorem 2 shows that the aggregated mean squared residual error over a large number of non-overlapping blocks converges at a higher rate, $\max(p^{-1}, n^{-1/2})$. This is intuitively interpretable since each block almost independently contains the same information on r_0 due to the approximate martingale difference structure of $V(r, \hat{F}_k^r) - V(r, H_k^r F_k^0)$ ($k = 1, \ldots, \lfloor n/k_n \rfloor$). This is clearly seen in the proof of (A10).

Remark 3. There are three main sources of error resulting in the temporal rate, $n^{-1/2}$, in the first two equalities of Theorem 2. They are the discretization error, the approximation error on the eigenvalue and eigenvector of $(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p$, and the aggregation efficiency. The orders of their convergence rates are k_n/n, $1/k_n$ and k_n/n, respectively, obtained by checking the proof of Theorem 2. This explains why $k_n = \lfloor\theta\sqrt{n}\rfloor$ is the best choice. Choosing an optimal θ for finite n is difficult in theory. One can choose one k_n from the candidate set in which the estimates of r_0 are stable. Our simulation experience tells us that it is always safe to choose θ such that $k_n \doteq \lfloor\sqrt{n}\rfloor$. The simulation results given in the Supplementary Material show that the estimates are not sensitive to θ when θ is around 1.

THEOREM 3. *Under Assumptions 1–3, if* $k_n = \lfloor\theta\sqrt{n}\rfloor$ *and* $g(p,n)$ *satisfies* $C_{pn}^2 g(p,n) \to \infty$ *and* $g(p,n) \to 0$ *as* $p,n \to \infty$, *then* $\lim_{p,n\to\infty} \mathrm{pr}(\hat{r}_0 = r_0) = 1$.

Theorem 3 demonstrates that one can choose a correct number of common factors, with probability approaching 1, using the method developed in § 2.

Remark 4. An example of $g(p,n)$ satisfying the conditions in Theorem 3 is $g(p,n) = (p + k_n)(pk_n)^{-1}\log\{pk_n(p+k_n)^{-1}\}$. This penalty function will be used in the numerical studies that follow.

5. NUMERICAL STUDIES

5·1. *Simulation studies*

In this section, we conduct simulations to check the theory for our estimator. We generate the data from stochastic volatility models of the form (2) with $r_0 = 3$ or 5. The instantaneous volatility processes σ_{it}^r $(i = 1,\ldots,p;\ r = 1,\ldots,r_0)$ are generated independently from the following square-root processes:

$$\mathrm{d}(\sigma_{it}^r)^2 = b_{ri}\{a_{ri} - (\sigma_{it}^r)^2\}\,\mathrm{d}t + \sigma_{ri}^0|\sigma_{it}^r|\,\mathrm{d}W_{it}^{\sigma}. \tag{9}$$

We set $a_{1i} = 0{\cdot}5 + i/p$, $a_{2i} = 0{\cdot}75 + i/p$, $a_{3i} = 0{\cdot}6 + i/p$, $a_{4i} = 0{\cdot}8 + i/p$, $a_{5i} = 0{\cdot}9 + i/p$, $b_{1i} = 0{\cdot}03 + i/(100p)$, $b_{2i} = 0{\cdot}05 + i/(100p)$, $b_{3i} = b_{4i} = b_{5i} = 0{\cdot}08 + i/(100p)$, $\sigma_{1i}^0 = 0{\cdot}15 + i/(10p)$, and $\sigma_{2i}^0 = \sigma_{3i}^0 = \sigma_{4i}^0 = \sigma_{5i}^0 = 0{\cdot}2 + i/(10p)$. These choices of parameter values are similar to those in Jacod & Todorov (2014) and Kong et al. (2015). The specific volatility process follows the stochastic differential equation

$$\mathrm{d}(\sigma_{it}^*)^2 = \{0{\cdot}08 + i/(100p)\}\{0{\cdot}25 + i/p - (\sigma_{it}^*)^2\}\,\mathrm{d}t + \{0{\cdot}2 + i/(10p)\}|\sigma_{it}^*|\,\mathrm{d}W_{it}^{\sigma *}.$$

The simulations are repeated 1000 times, with r_m set to 6 and $\beta = \hat{\xi} = U(6, \hat{F}^6)$. We use the penalty function presented in Remark 4.

We first consider the case where $n = 1560$, mimicking 20 days of data taken every five minutes. We tune θ in the sense of Remark 3, such that $k_n = 39 \doteq \lfloor\sqrt{1560}\rfloor$ and hence the dataset is split into 40 blocks. Next, we consider the case where $n = 780$, mimicking 10 days of data taken every five minutes. Again, we tune θ so that $k_n = 30 \doteq \lfloor\sqrt{780}\rfloor$ and hence the dataset is conveniently split into 26 blocks. The simulation results for $n = 1560$ and $n = 780$ are reported in the left columns of Table 1. From the left five columns of Table 1 we make the following observations. The estimates of common factors are close to the true values when p is large. For fixed n, as p

Table 1. *Estimates of the number of factors, with standard errors in parentheses*

	LPCA				GE1		GE2	
	$n = 1560$		$n = 780$		$n = 780$		$n = 780$	
p	$r_0 = 3$	$r_0 = 5$	$r_0 = 3$	$r_0 = 5$	$r_0 = 3$	$r_0 = 5$	$r_0 = 3$	$r_0 = 5$
40	4·9 (0·26)	5·8 (0·37)	5·0 (0·10)	5·8 (0·37)	5·5 (0·53)	6 (0)	5·0 (0·20)	5·0 (0·11)
60	4·0 (0·10)	5·0 (0·18)	4·1 (0·12)	5·0 (0·11)	5·8 (0·41)	6 (0)	5·0 (0·14)	5·0 (0·08)
80	3·2 (0·49)	5 (0)	3·6 (0·50)	5 (0)	6 (0·10)	6 (0)	4·9 (0·28)	5 (0)
100	3 (0)	5 (0)	3·0 (0·10)	5 (0)	6 (0·05)	6 (0)	4·9 (0·24)	5 (0)
120	3 (0)	5 (0)	3 (0)	5 (0)	6 (0)	6 (0)	4·9 (0·22)	5 (0)
140	3 (0)	5 (0)	3 (0)	5 (0)	6 (0)	6 (0)	4·9 (0·18)	5 (0)
160	3 (0)	5 (0)	3 (0)	5 (0)	6 (0)	6 (0)	5 (0·15)	5 (0)
180	3 (0)	5 (0)	3 (0)	5 (0)	6 (0)	6 (0)	5 (0·15)	5 (0)
200	3 (0)	5 (0)	3 (0)	5 (0)	6 (0)	6 (0)	5 (0·14)	5 (0)

LPCA, estimate based on local principal component analysis; GE1, the first global estimate; GE2, the second global estimate.

increases, the estimate becomes more accurate; in particular, it works almost perfectly when p is greater than 100. Overall, the estimates for $n = 780$ are less accurate than those for $n = 1560$ because of the smaller sample size. This is consistent with Theorem 2, which states that the aggregated mean squared residual error converges more quickly as n or p increases, resulting in more accurate and stable estimates of r_0.

We further compare two estimators based on the global principal component analysis. We consider only the case where $n = 780$; the results for $n = 1560$ are similar and therefore omitted for conciseness. The first global estimator is obtained by directly applying the method of Bai & Ng (2002) to the full range of the high-frequency data or, equivalently, setting $k_n = 780$. The second estimator is the one presented in Aït-Sahalia & Xiu (2017), with the same tuning parameters as in their simulation studies. The results for the two global estimates are given in the right four columns of Table 1.

Table 1 shows that the global principal component analysis estimator works well for $r_0 = 5$ but has an intrinsic bias when $r_0 = 3$. The first global estimate is unacceptable for both values of r_0. Both methods overestimate the number of common factors, due to the time-varying feature of the factor loading process. The global methods suffer from a large discretization error that cannot be eliminated by increasing p, which can be seen from the right columns of Table 1. This demonstrates the superiority of the aggregation of local factor analysis.

Additional simulation studies are reported in the Supplementary Material. The results demonstrate that our estimator is not sensitive to the nonstationarity of the volatility processes, to the tuning parameter θ when it is around 1, or to the microstructure noise when the sampling length is up to the minute scale.

5·2. *Real-data analysis*

In this section, we implement our approach by analysing data on the 99 most heavily traded stocks constituting the S&P 500 index in April 2013. Collecting data on the most liquid stocks helps to reduce synchronization bias, because the equally spaced observations may not be trading prices but rather the previous ticks immediately before the observation times. In order to avoid the adverse effects of microstructure noise and asynchronicity bias, we sample the dataset sparsely. We collect $n = 780$ observations of five-minute or ten-minute returns for these 99 stocks. To avoid overnight jumps and the weekend effect, overnight and weekend returns are excluded. Motivated

by Remark 3 and consistent with the simulation studies, we choose $k_n = 30$. We first set $r_m = 10$. We estimate the average variance of the idiosyncratic component by $U(4, \hat{F}^4)$, assuming that r_0 is close to four, a prior number implied by the Fama–French four-factor model and many other empirical studies.

When the five-minute returns starting at 9:30 AM on 1 April are used, our estimated number of common factors is five. To check the stability of the procedure, for every ten-minute interval ahead of the last starting time until 11:20 AM on 2 April, we further collect 780 five-minute returns and calculate the estimates of r_0. The results show that all estimates equal 5. Similar findings are obtained when we change r_m to 8 and to 6. This demonstrates that the number of common factors is stable over time and robust with respect to the choice of r_m.

To observe the results when we sample the data more sparsely, we use ten-minute returns with different starting minutes. As in the five-minute case, all estimates equal 5. This demonstrates the stability of our estimator across sampling frequencies. Next, to check the stability cross-sectionally, we separate the stocks into two groups according to their natural location, with 49 stocks in the first group and the others in the second group. For both five-minute and ten-minute returns, the estimates in the first group and the second group are the same, with all being equal to 5.

To illustrate an estimation of the number of factors by minimizing the aggregated mean squared residual error, a graph of the error against r is plotted in the Supplementary Material.

6. Discussion

This work provides a basic tool for the further analysis of high-dimensional high-frequency data. Some interesting future research problems are the following.

Once r_0 in model (2) is determined, it is possible to estimate the high-dimensional integrated volatility matrix using the factor-based approach. Wang & Zou (2010) were the first to develop a thresholding estimator of the large integrated volatility matrix. Applied to the factor-corrected data, their method can be used to estimate the large integrated specific volatility matrix.

Some practical issues must be studied carefully using ultrahigh-frequency data. What is the effect of the microstructure noise and how can it be eliminated? How large is the bias caused by asynchronicity and how can we correct for it? When working with high- but not ultrahigh-frequency data, our method can be extended to include jumps in driving forces and volatility processes in model (2) using more complicated mathematics. Separating the continuous driving forces and jump driving forces is of considerable interest, and we leave these problems to future research.

Acknowledgement

The author's work was supported in part by the Natural Science Foundation of China, the Humanities and Social Science Youth Foundation of the Chinese Ministry of Education, and the Priority Academic Program Development project of Jiangsu higher education institutions.

Supplementary material

Supplementary material available at *Biometrika* online includes additional numerical studies and proofs of Lemmas A1–A6, which are technical results supporting the proofs of the main theorems.

APPENDIX

Technical lemmas

In the proofs, C will stand for a generic positive constant that may take different values in different places. For simplicity, let $\mathcal{F}_{k-1} = \mathcal{F}_{t_{(k-1)k_n}}$. By the standard localization method, it suffices to prove the results under the following strengthened assumption.

Assumption A1. We have

$$\max_{1\leqslant i\leqslant p}\sup_{0\leqslant t\leqslant T}(|\mu_{it}|+|\sigma^r_{it}|+|\sigma^*_{it}|)<C,\quad \inf_{0\leqslant t\leqslant T}\lambda_{\min}(\sigma_t^{\mathrm{T}}\sigma_t/p)\geqslant C^{-1},$$
$$\max_{1\leqslant i\leqslant p}\{|E_{\mathcal{F}_t}(Z_{t+s}-Z_t)|+E_{\mathcal{F}_t}(Z_{t+s}-Z_t)^2\}\leqslant Cs,\quad Z=\sigma_i^r,\sigma_i^*.$$

The distances between adjacent eigenvalues of $\sigma_t^{\mathrm{T}}\sigma_t/p$ are bounded away from zero for all $0\leqslant t\leqslant T$.

We first present some technical lemmas, the proofs of which are given in the Supplementary Material.

LEMMA A1. *If Assumptions 1–3 and A1 hold, then for $l=1,2,4$,*

$$E_{\mathcal{F}_{k-1}}\{\|(\sigma_k^0F_k^0)^{\mathrm{T}}\sigma_k^0F_k^0\|^l(pk_n)^{-l}\}\leqslant C,\quad \max_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\|(\sigma_k^0F_k^0)^{\mathrm{T}}\sigma_k^0F_k^0\|(pk_n)^{-1}=O_{\mathrm{p}}(1),\tag{A1}$$
$$E_{\mathcal{F}_{k-1}}\{(pk_n)^{-2}\|(\sigma_k^*F_k^*)^{\mathrm{T}}\sigma_k^*F_k^*\|^2\}\leqslant CC_{pn}^{-2},\tag{A2}$$
$$\max_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\|(F_k^*)^{\mathrm{T}}F_k^*-E\{(F_k^*)^{\mathrm{T}}F_k^*\}\|(pk_n)^{-1}=o_{\mathrm{p}}(1).\tag{A3}$$

LEMMA A2. *Let $\hat{D}_k=k_n\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}$ and $D_k=k_n\{H_k^rF_k^0(H_k^rF_k^0)^{\mathrm{T}}\}^{-1}$. If Assumptions 1–3 and A1 hold, then for $l=1,2,4$ we have*

$$E_{\mathcal{F}_{k-1}}\left[\left\{k_n^{-1}\sum_{j=1}^{k_n}\|\hat{F}_k^r(j)-H_k^rF_k^0(j)\|^2\right\}^l\right]\leqslant CC_{pn}^{-2l},\tag{A4}$$
$$E_{\mathcal{F}_{k-1}}\{\|\hat{D}_k^{-1}-D_k^{-1}\|^l\}\leqslant CC_{pn}^{-l},\quad \max_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\|\hat{D}_k^{-1}-D_k^{-1}\|=o_{\mathrm{p}}(1).\tag{A5}$$

LEMMA A3. *If Assumptions 1–3 and A1 hold, then*

$$\mathrm{pr}\left\{C<\min_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\lambda_{\min}(D_k^{-1})\leqslant\max_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\lambda_{\max}(D_k^{-1})<C^{-1}\right\}=1+o(1),\tag{A6}$$
$$\mathrm{pr}\left\{C<\min_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\lambda_{\min}(\hat{D}_k^{-1})\leqslant\max_{1\leqslant k\leqslant\lfloor n/k_n\rfloor}\lambda_{\max}(\hat{D}_k^{-1})<C^{-1}\right\}=1+o(1).\tag{A7}$$

LEMMA A4. *If Assumptions 1–3 and A1 hold, then*

$$(\lfloor n/k_n\rfloor pk_n^2)^{-1}\sum_{k=1}^{\lfloor n/k_n\rfloor}\sum_{i=1}^{p}\delta_k^{\mathrm{T}}(i)(\hat{F}_k^r-H_k^rF_k^0)^{\mathrm{T}}\hat{D}_k(\hat{F}_k^r-H_k^rF_k^0)\delta_k(i)=O_{\mathrm{p}}(C_{pn}^{-2}),\tag{A8}$$

$$(\lfloor n/k_n \rfloor p k_n^2)^{-1} \sum_{k=1}^{\lfloor n/k_n \rfloor} \sum_{i=1}^{p} \delta_k^{\mathrm{T}}(i)(H_k^r F_k^0)^{\mathrm{T}}(\hat{D}_k - D_k) H_k^r F_k^0 = O_{\mathrm{p}}(C_{pn}^{-2}), \tag{A9}$$

$$(\lfloor n/k_n \rfloor p k_n^2)^{-1} \sum_{k=1}^{\lfloor n/k_n \rfloor} \sum_{i=1}^{p} \delta_k^{\mathrm{T}}(i)(\hat{F}_k^r - H_k^r F_k^0)^{\mathrm{T}} \hat{D}_k H_k^r F_k^0 = O_{\mathrm{p}}(C_{pn}^{-2}). \tag{A10}$$

LEMMA A5. *Let* $\tilde{D}_k = (F_k^0)^{\mathrm{T}}\{F_k^0(F_k^0)^{\mathrm{T}}\}^{-1}F_k^0 - (H_k^r F_k^0)^{\mathrm{T}}\{H_k^r F_k^0 (H_k^r F_k^0)^{\mathrm{T}}\}^{-1} H_k^r F_k^0$*; then under Assumptions 1–3 and A1,*

$$\sum_{k=1}^{\lfloor n/k_n \rfloor} \mathrm{tr}\{\sigma_k^* F_k^* \tilde{D}_k (\sigma_k^0 F_k^0)^{\mathrm{T}} + \sigma_k^0 F_k^0 \tilde{D}_k (\sigma_k^* F_k^*)^{\mathrm{T}}\}/(\lfloor n/k_n \rfloor p k_n) = O_{\mathrm{p}}(C_{pn}^{-1}),$$

$$\sum_{k=1}^{\lfloor n/k_n \rfloor} [\mathrm{tr}\{\sigma_k^0 F_k^0 \tilde{D}_k (\sigma_k^0 F_k^0)^{\mathrm{T}}\}/(p k_n) + \mathrm{tr}\{\bar{\delta}_k^{\mathrm{T}} \tilde{D}_k (\delta_k - \bar{\delta}_k)\}/(\lfloor n/k_n \rfloor p k_n)] = O_{\mathrm{p}}(C_{pn}^{-1}),$$

$$\sum_{k=1}^{\lfloor n/k_n \rfloor} [\mathrm{tr}\{\sigma_k^* F_k^* \tilde{D}_k (F_k^* \sigma_k^*)^{\mathrm{T}}\}/(p k_n) + \mathrm{tr}\{(\delta_k - \bar{\delta}_k)^{\mathrm{T}} \tilde{D}_k \delta_k\}/(\lfloor n/k_n \rfloor p k_n)] = O_{\mathrm{p}}(C_{pn}^{-1}).$$

LEMMA A6. *Suppose that Assumptions 1–3 and A1 hold. Let*

$$\begin{aligned} V_k'(2) &= 2(pk_n)^{-1} \mathrm{tr}[\sigma_k^0 H_k^{r-}(\hat{F}_k^r - H_k^r F_k^0)(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r(\sigma_k^* F_k^*)^{\mathrm{T}}] \\ &\quad - 2(pk_n)^{-1} \mathrm{tr}\{\sigma_k^0 H_k^{r-}(\hat{F}_k^r - H_k^r F_k^0)(\sigma_k^* F_k^*)^{\mathrm{T}}\}, \\ V_k'(3) &= -(pk_n)^{-1} \mathrm{tr}[\sigma_k^0 H_k^{r-}(\hat{F}_k^r - H_k^r F_k^0)(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r(\hat{F}_k^r - H_k^r F_k^0)^{\mathrm{T}}(\sigma_k^0 H_k^{r-})^{\mathrm{T}}] \\ &\quad + (pk_n)^{-1} \mathrm{tr}\{\sigma_k^0 H_k^{r-}(\hat{F}_k^r - H_k^r F_k^0)(\hat{F}_k^r - H_k^r F_k^0)^{\mathrm{T}}(\sigma_k^0 H_k^{r-})^{\mathrm{T}}\}. \end{aligned}$$

Then $\lfloor n/k_n \rfloor^{-1} \sum_{k=1}^{\lfloor n/k_n \rfloor}\{V_k'(2) + V_k'(3)\} = O_{\mathrm{p}}(C_{pn}^{-2})$.

Proof of Theorem 1

We first prove (6). Let L_{nk} be the diagonal matrix consisting of the r largest eigenvalues of $\delta\delta^{\mathrm{T}}/(pk_n)$. Then we have $k_n^{-1}\bar{F}_k^r \delta_k \delta_k^{\mathrm{T}}(pk_n)^{-1}(\bar{F}_k^r)^{\mathrm{T}} = L_{nk}$. By Weyl's theorem, (A2), and inequalities (4), (20) and (21) in the Supplementary Material, it suffices to prove that $E_{\mathcal{F}_{k-1}}\{\|(F_k^0)^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}\sigma_k^* F_k^*\|^2 (pk_n)^{-2}\} \leqslant CC_{pn}^{-2}$, which is a consequence of Assumption 2 and

$$\begin{aligned} (pk_n)^{-2}\|F_k^{0\prime}\sigma_k^{0\prime}\sigma_k^* F_k^*\|^2 &\leqslant \mathrm{tr}\{(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\sigma_k^0 F_k^0 (F_k^0)^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}\sigma_k^* F_k^* (pk_n)^{-2}\} \\ &= p^{-2}k_n^{-2} \sum_{m=1}^{k_n}\sum_{l=1}^{k_n} \left\{\sum_{i=1}^{p} F_k^*(i,j)\sigma_k^*(i,i) \sum_{l=1}^{r} \sigma_k^0(i,l) F_k^0(l,m)\right\}^2. \end{aligned} \tag{A11}$$

Next, we prove (7). Let

$$\begin{aligned} d_k &= \{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}^{1/2}(pk_n)^{-1}[F_k^0(F_k^0)^{\mathrm{T}}k_n^{-1}(\sigma_k^0)^{\mathrm{T}}\{(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\}^{\mathrm{T}}(\bar{F}_k^r)^{\mathrm{T}} \\ &\quad + F_k^0(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\sigma_k^0(F_k^0)^{\mathrm{T}}(\bar{F}_k^r)^{\mathrm{T}}k_n^{-1} + F_k^0(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\sigma_k^* F_k^*(\bar{F}_k^r)^{\mathrm{T}}k_n^{-1}] \end{aligned}$$

and $\gamma_{nk} = R_k\{\mathrm{diag}(R_k^{\mathrm{T}}R_k)\}^{-1/2}$, where $R_k = \{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}^{1/2}F_k^0(\bar{F}_k^r)^{\mathrm{T}}/k_n$.

By the relation $\delta_k\delta_k^{\mathrm{T}}(pk_n)^{-1}(\bar{F}_k^r)^{\mathrm{T}} = (\bar{F}_k^r)^{\mathrm{T}}L_{nk}$, we have $(\tilde{B}_k + d_k R_k^{-1})\gamma_{nk} = \gamma_{nk}L_{nk}$, where $\tilde{B}_k = \{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}^{1/2}\{F_k^0(F_k^0)^{\mathrm{T}}/k_n\}\{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}^{1/2}$. By Assumptions 2 and 3 and a similar proof to the one for (A6) we have, with probability approaching 1,

$$\|R_k^{-1}\| = \lambda_{\max}^{1/2}\{(R_k^{-1})^{\mathrm{T}}R_k^{-1}\} = \lambda_{\min}^{-1/2}\{R_k R_k^{\mathrm{T}}\} \leqslant \lambda_{\min}^{-1/2}\{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}\lambda_{\min}^{-1/2}\{F_k^0(F_k^0)^{\mathrm{T}}/k_n\}, \tag{A12}$$

which is finite uniformly in $k = 1, \ldots, \lfloor n/k_n \rfloor$. Denote $\{\max_{1 \leqslant k \leqslant \lfloor n/k_n \rfloor} \|R_k^{-1}\| < M\}$ by $\mathcal{A}$. By Assumption A1, (A1), (A2), (A11) and equation (2) in the Supplementary Material, we have $E_{\mathcal{F}_{k-1}}(\|d_k\|) \leqslant C/C_{pn}$. This, together with (A12) and Theorem 2 in Davis & Kahan (1970), shows that $E_{\mathcal{F}_{k-1}}(\|\gamma_{nk} - \gamma_k\|)I(\mathcal{A}) \leqslant C/C_{pn}$. Now, by the definition of R_k, we get $F_k^0(\bar{F}_k^r)^{\mathrm{T}}/k_n = \{(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p\}^{-1/2}\gamma_{nk}L_{nk}^*$, where $L_{nk}^* = \{\mathrm{diag}(R_k^{\mathrm{T}}R_k)\}^{1/2}$. Notice that $\|\gamma_{nk}\| = 1$ and $\max_k \|L_k\| < \infty$; then (7) is readily obtained from (6) and the fact that $P(\mathcal{A}^{\mathrm{c}}) = o(1)$. Finally, we prove the third equation in the theorem. Let $(e_k^r)^{\mathrm{T}}$, $\bar{e}_k^r$ and $\tilde{e}_k^r$ be the eigenvector matrices of $(F_k^0)^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}\sigma_k^0F_k^0/(pk_n)$, $(\sigma_k^0)^{\mathrm{T}}\sigma_k^0F_k^0(F_k^0)^{\mathrm{T}}/(pk_n)$ and $(\sigma_k^0)^{\mathrm{T}}\sigma_k^0/p$, respectively. Then, by Theorem 2 in Davis & Kahan (1970), we have $E_{\mathcal{F}_{k-1}}(\|\bar{F}_k^r/\sqrt{k_n} - e_k^r\|) \leqslant C/C_{pn}$ and $E_{\mathcal{F}_{k-1}}(\|\bar{e}_k^r - \tilde{e}_k^r\|) \leqslant C/C_{pn}$. On the other hand, the ith row of e_k^r, denoted by $(e_{ki}^r)^{\mathrm{T}}$ $(i = 1, \ldots, r)$, satisfies $e_{ki}^r = \{\bar{e}_k^r(i)\}^{\mathrm{T}}F_k^0/\|\{\bar{e}_k^r(i)\}^{\mathrm{T}}F_k^0\|$, where $\bar{e}_k^r(i)$ is the ith column of $\bar{e}_k^r$. This completes the proof.

Proof of Theorem 2

Let $P_k = k_n^{-1}(F_k^0)^{\mathrm{T}}(H_k^r)^{\mathrm{T}}D_kH_k^rF_k^0$ and $\hat{P}_k = k_n^{-1}(\hat{F}_k^r)^{\mathrm{T}}\hat{D}_k\hat{F}_k^r$. Simple algebraic manipulation yields

$$\begin{aligned}\hat{P}_k - P_k = k_n^{-1}\big[&\{(\hat{F}_k^r)^{\mathrm{T}} - (F_k^0)^{\mathrm{T}}(H_k^r)^{\mathrm{T}}\}\hat{D}_k(\hat{F}_k^r - H_k^rF_k^0) + \{(\hat{F}_k^r)^{\mathrm{T}} - (F_k^0)^{\mathrm{T}}(H_k^r)^{\mathrm{T}}\hat{D}_kH_k^rF_k^0\\ &+ (F_k^0)^{\mathrm{T}}(H_k^r)^{\mathrm{T}}\hat{D}_k(\hat{F}_k^r - H_k^rF_k^0) + (F_k^0)^{\mathrm{T}}(H_k^r)^{\mathrm{T}}(\hat{D}_k - D_k)H_k^rF_k^0\big],\end{aligned}$$

Notice that $V(r, \hat{F}_k^r) - V(r, H_k^rF_k^0) = (pk_n)^{-1}\sum_{i=1}^p \delta_k^{\mathrm{T}}(i)(P_k - \hat{P}_k)\delta_k(i)$, where $\delta_k(i)$ is the ith column of δ_k. This, together with Lemma A4, proves the result for $r \leqslant r_0$. The result for $r < r_0$ is a consequence of Lemma A5 and the following decomposition:

$$\begin{aligned}V(r, H_k^rF_k^0) - V(r_0, F_k^0) = &\,\mathrm{tr}\{\sigma_k^*F_k^*\tilde{D}_k(F_k^0)^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}} + \sigma_k^0F_k^0\tilde{D}_k(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\}/(pk_n)\\ &+ \mathrm{tr}\{\sigma_k^0F_k^0\tilde{D}_k(F_k^0)^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}\}/(pk_n) + \mathrm{tr}\{\bar{\delta}_k^{\mathrm{T}}\tilde{D}_k(\delta_k - \bar{\delta}_k)\}/(pk_n)\\ &+ \mathrm{tr}\{\sigma_k^*F_k^*\tilde{D}_k(\sigma_k^*)^{\mathrm{T}}(F_k^*)^{\mathrm{T}}\}/(pk_n) + \mathrm{tr}\{(\delta_k - \bar{\delta}_k)^{\mathrm{T}}\tilde{D}_k\delta_k\}/(pk_n).\end{aligned}$$

Next, we prove the result for $r > r_0$. Since $r > r_0$, H_k^r is of rank r_0. Let H_k^{r-} be the generalized inverse of H_k^r, such that $H_k^{r-}H_k^r = I_r$; then $\bar{\delta}_k^{\mathrm{T}} = \sigma_k^0H_k^{r-}H_k^rF_k^0 + \sigma_k^*F_k^*$. Now we have the decomposition

$$\begin{aligned}&V(r, \hat{F}_k^r) - V(r_0, F_k^0)\\ &\quad= (pk_n)^{-1}\,\mathrm{tr}[\sigma_k^*F_k^*\{(F_k^0)^{\mathrm{T}}\{F_k^0(F_k^0)^{\mathrm{T}}\}^{-1}F_k^0 - (\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r\}(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}]\\ &\qquad+ 2(pk_n)^{-1}\,\mathrm{tr}[\sigma_k^0H_k^{r-}(\hat{F}_k^r - H_k^rF_k^0)(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}]\\ &\qquad- 2(pk_n)^{-1}\,\mathrm{tr}\{\sigma_k^0H_k^{r-}(\hat{F}_k^r - H_k^rF_k^0)(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\}\\ &\qquad- (pk_n)^{-1}\,\mathrm{tr}[\sigma_k^0H_k^{r-}(\hat{F}_k^r - H_k^rF_k^0)(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r(\hat{F}_k^r - H_k^rF_k^0)^{\mathrm{T}}(H_k^{r-})^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}]\\ &\qquad+ (pk_n)^{-1}\,\mathrm{tr}\{\sigma_k^0H_k^{r-}(\hat{F}_k^r - H_k^rF_k^0)(\hat{F}_k^r - H_k^rF_k^0)^{\mathrm{T}}(H_k^{r-})^{\mathrm{T}}(\sigma_k^0)^{\mathrm{T}}\}\\ &\quad= V_k'(1) + V_k'(2) + V_k'(3),\end{aligned}$$

where $V_k'(1)$ is the first term, $V_k'(2)$ is the sum of the second and third terms, and $V_k'(3)$ is the sum of the fourth and fifth terms. Define $V_k'(1) = V_k'(1,2) - V_k'(1,1)$. By the decomposition and Lemma A6, it is enough to prove that $\lfloor n/k_n \rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n \rfloor} V_k'(1,1) = O_{\mathrm{p}}(C_{pn}^{-2})$. We separate the proof into two steps. In the first step, we show that $\lfloor n/k_n \rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n \rfloor}\mathrm{tr}(\underline{D}_k)/(pk_n) = O_{\mathrm{p}}(C_{pn}^{-2})$, where $\underline{D}_k = \sigma_k^*F_k^*[(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r - (\tilde{F}_k^r)^{\mathrm{T}}\{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}\tilde{F}_k^r](\sigma_k^*F_k^*)^{\mathrm{T}}$ and $\tilde{F}_k^r = \breve{F}_k^r\bar{\delta}_k\bar{\delta}_k^{\mathrm{T}}/(pk_n)$, with $\breve{F}_k^r$ being $\sqrt{k_n}$ times the $r \times k_n$ matrix of eigenvectors listed in the rows of $\bar{\delta}_k\bar{\delta}_k^{\mathrm{T}}/(pk_n)$. Since $\bar{\delta}_k$ has constant factor loading matrix conditioning on $\mathcal{F}_{k-1}$, by the argument given in the errata document at http://www.columbia.edu/~sn2294/papers/correctionEcta2.pdf, in the second step we have

$$\lfloor n/k_n \rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n \rfloor}(pk_n)^{-1}\,\mathrm{tr}(\sigma_k^*F_k^*[(\tilde{F}_k^r)^{\mathrm{T}}\{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}\tilde{F}_k^r](\sigma_k^*F_k^*)^{\mathrm{T}}) = O_{\mathrm{p}}(C_{pn}^{-2}).$$

Now we return to the proof of the first step. We have the decomposition

$$\begin{aligned}(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r - \tilde{F}_k^{r\prime}\{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}\tilde{F}_k^r &= (\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}(\hat{F}_k^r - \tilde{F}_k^r)\\ &\quad + (\hat{F}_k^r)^{\mathrm{T}}[\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1} - \{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}]\tilde{F}_k^r\\ &\quad + \{(\hat{F}_k^r)^{\mathrm{T}} - (\tilde{F}_k^r)^{\mathrm{T}}\}\{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}\tilde{F}_k^r.\end{aligned}$$

Because the rank of $\hat{F}_k^r$ is no larger than r,

$$\begin{aligned}&(pk_n)^{-1}\big|\mathrm{tr}[\sigma_k^*F_k^*(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}(\hat{F}_k^r - \tilde{F}_k^r)(F_k^*)^{\mathrm{T}}\sigma_k^*]\big|\\ &\quad\leqslant (pk_n)^{-1/2}(\mathrm{tr}[\sigma_k^*F_k^*(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\hat{F}_k^r(\sigma_k^*F_k^*)^{\mathrm{T}}])^{1/2}\\ &\qquad\times (pk_n)^{-1/2}(\mathrm{tr}[\sigma_k^*F_k^*\{\hat{F}_k^r - \tilde{F}_k^r\}^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\{\hat{F}_k^r - \tilde{F}_k^r\}(\sigma_k^*F_k^*)^{\mathrm{T}}])^{1/2}\\ &\quad\leqslant r_m\|(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\sigma_k^*F_k^*\|/(pk_n)\|\{\hat{F}_k^r - \tilde{F}_k^r\}^{\mathrm{T}}/k_n^{1/2}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}/k_n\}^{-1}(\hat{F}_k^r - \tilde{F}_k^r)/k_n^{1/2}\|^{1/2}\\ &\quad\leqslant r_m\|(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}\sigma_k^*F_k^*\|/(pk_n)\|(\hat{F}_k^r - \tilde{F}_k^r)(\hat{F}_k^r - \tilde{F}_k^r)^{\mathrm{T}}/k_n\|^{1/2}\|\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}k_n\|^{1/2}. \qquad \text{(A13)}\end{aligned}$$

Inequality (21) in the Supplementary Material shows that $E_{\mathcal{F}_{k-1}}\{\|(\hat{F}_k^r - \tilde{F}_k^r)(\hat{F}_k^r - \tilde{F}_k^r)^{\mathrm{T}}/k_n\|\} \leqslant C/C_{pn}^2$. This, together with (A13) and Lemmas A1 and A3, implies that

$$\lfloor n/k_n\rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n\rfloor}(pk_n)^{-1}\big|\mathrm{tr}[\sigma_k^*F_k^*(\hat{F}_k^r)^{\mathrm{T}}\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1}\{\hat{F}_k^r - \tilde{F}_k^r\}(F_k^*)^{\mathrm{T}}(\sigma_k^*)^{\mathrm{T}}]\big| = O_{\mathrm{p}}(C_{pn}^{-2}).$$

Similarly, we obtain the two equations

$$\begin{aligned}&\lfloor n/k_n\rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n\rfloor}(pk_n)^{-1}\big|\mathrm{tr}[\sigma_k^*F_k^*(\hat{F}_k^r - \tilde{F}_k^r)^{\mathrm{T}}\{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}\tilde{F}_k^rF_k^{*\prime}\sigma_k^{*\prime}]\big| = O_{\mathrm{p}}(C_{pn}^{-2}),\\ &\lfloor n/k_n\rfloor^{-1}\sum_{k=1}^{\lfloor n/k_n\rfloor}(pk_n)^{-1}\big|\mathrm{tr}(\sigma_k^*F_k^*(\hat{F}_k^r)^{\mathrm{T}}[\{\hat{F}_k^r(\hat{F}_k^r)^{\mathrm{T}}\}^{-1} - \{\tilde{F}_k^r(\tilde{F}_k^r)^{\mathrm{T}}\}^{-1}]\tilde{F}_k^r(\sigma_k^*F_k^*)^{\mathrm{T}})\big| = O_{\mathrm{p}}(C_{pn}^{-2}).\end{aligned}$$

The above three equations finish the proof of the first step.

Proof of Theorem 3

It suffices to show that $\mathrm{pr}\{U_p(r) - U_p(r_0) < 0\} = o(1)$. Since $H_k^{r_0}F_k^0$ and F_k^0 span the same column space, $U(r_0, H^{r_0}\circ F^0) - U(r_0, F^0) = 0$. Then

$$U_p(r) - U_p(r_0) = U(r,\hat{F}^r) - U(r, H^r\circ F^0) + U(r, H^r\circ F^0) - U(r_0, H^{r_0}\circ F^0) - (r_0 - r)g(p,n).$$

For $r < r_0$, by the first and third equations of Theorem 2, $U(r,\hat{F}^r) - U(r, H^r\circ F^0) = O_{\mathrm{p}}(C_{pn}^{-2})$ while $U(r, H^r\circ F^0) - U(r_0, H^{r_0}\circ F^0)$ has a positive lower bound in probability. This proves the theorem since $g(p,n) \to 0$ as $p, n \to \infty$. Consider $r \geqslant r_0$. By the second equation of Theorem 2, we have $U(r,\hat{F}^r) - U(r_0, F^0) = O_{\mathrm{p}}(C_{pn}^{-2})$. This proves the theorem since $C_{pn}^2 g(p,n) \to \infty$.

References

Ahn, S. & Hohenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–27.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Finan. Stud.* **18**, 351–416.

AÏT-SAHALIA, Y. & XIU, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high frequency. *J. Economet.* to appear.
ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. & LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625.
BAI, J. & NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.
BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. & SHEPHARD, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481–536.
BARNDORFF-NIELSEN, O. E. & SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Statist. Soc.* B **64**, 253–80.
CHAMBERLAIN, G. & ROTHSCHILD, M. (1983). Arbitarge, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–304.
CHRISTENSEN, K., KINNEBROCK, S. & PODOLSKIJ, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Economet.* **159**, 116–33.
DAVIS, C. & KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7**, 1–46.
ERBAN, R. & OTHMER, H. (2014). Editorial: Special issue on stochastic modeling of reaction-diffusion processes in biology. *Bull. Math. Biol.* **76**, 761–5.
FAMA, E. & FRENCH, K. (1992). The cross-section of expected stock returns. *J. Finance* **47**, 427–65.
FAN, J., FURGER, A. & XIU, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *J. Bus. Econ. Statist.* **34**, 489–503.
FAN, J., LI, Y. & KE, Y. (2012). Vast volatility matrix estimation using high frequency data for portfolio selection. *J. Am. Statist. Assoc.* **107**, 412–28.
FAN, J., LIAO, Y. & MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc.* B **75**, 603–80.
FLEMING, W. (1975). Diffusion processes in population biology. *Adv. Appl. Prob.* **7**, 100–5.
JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. & VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stoch. Proces. Appl.* **119**, 2249–76.
JACOD, J. & PODOLSKIJ, M. (2013). A test for the rank of the volatility process: The random perturbation approach. *Ann. Statist.* **41**, 2391–427.
JACOD, J. & ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41**, 1462–84.
JACOD, J. & TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42**, 1029–69.
JING, B. Y., KONG, X. B., LIU, Z. & MYKLAND, P. (2012a). On the jump activity index on semimartingales. *J. Economet.* **166**, 213–23.
JING, B. Y., KONG, X. B. & LIU, Z. (2012b). Modeling high frequency financial data by pure jump models. *Ann. Statist.* **40**, 759–84.
KONG, X. B., LIU, Z. & JING, B. Y. (2015). Testing for pure jump processes for high-frequency data. *Ann. Statist.* **43**, 847–77.
LI, J., TODOROV, V. & TAUCHEN, G. (2013). Volatility occupation times. *Ann. Statist.* **41**, 1865–91.
LI, Y. & MYKLAND, P. A. (2008). Are volatility estimators robust with respect to modeling assumptions? *Bernoulli* **13**, 601–22.
MYKLAND, P. A. & ZHANG, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist.* **34**, 1931–63.
MYKLAND, P. A. & ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77**, 1403–45.
ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise. *J. Economet.* **168**, 244–58.
STOCK, J. & WATSON, M. (2002). Forecasting using principal components from large number of predictors. *J. Am. Statist. Assoc.* **97**, 1167–79.
TODOROV, V. & TAUCHEN, G. (2012a). The realized Laplace transform of volatility. *Econometrica* **80**, 1105–27.
TODOROV, V. & TAUCHEN, G. (2012b). Inverse realized Laplace transform for nonparametric volatility density estimation in jump diffusions. *J. Am. Statist. Assoc.* **107**, 622–35.
WANG, Y. & ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **38**, 943–78.
ZHANG, L., MYKLAND, P. A. & AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Am. Statist. Assoc.* **100**, 1394–411.

[*Received on* 15 *July* 2016. *Editorial decision on* 12 *February* 2017]