Large-dimensional factor modeling based on high-frequency observations

Markus Pelger*

August 20, 2015

Abstract

This paper develops a statistical theory to estimate an unknown factor structure based on financial high-frequency data. I derive a new estimator for the number of factors and derive consistent and asymptotically mixed-normal estimators of the loadings and factors under the assumption of a large number of cross-sectional and high-frequency observations. The estimation approach can separate factors for normal "continuous" and rare jump risk. The estimators for the loadings and factors are based on the principal component analysis of the quadratic covariation matrix. The estimator for the number of factors uses a perturbed eigenvalue ratio statistic. The results are obtained under general conditions, that allow for a very rich class of stochastic processes and for serial and cross-sectional correlation in the idiosyncratic components.

Keywords: Systematic risk, High-dimensional data, High-frequency data, Latent factor model, PCA, Jumps, Semimartingales, Approximate factor model, Number of factors

JEL classification: C14, C38, C55, C58

*Department of Management Science & Engineering, Stanford University, Stanford, CA 94305, Email: mpelger@stanford.edu. This paper is part of my Ph.D. thesis in Economics at UC Berkeley. I thank Robert M. Anderson, Martin Lettau, Michael Jansson and Lisa Goldberg for extensive advice and support. I also thank Jason Zhu for excellent research assistance. I am very grateful for comments and helpful discussions with Richard Stanton, Ulrike Malmendier, Steve Evans, Nicolae Gârleanu, Johan Walden, Noureddine El Karoui, Viktor Todorov and the participants at the Risk Management Seminar, Econometrics Seminar and Finance Pre-Seminar at UC Berkeley and the INFORMS Annual Meeting. This work was supported by the Center for Risk Management Research at UC Berkeley.

1 Introduction

Financial economists are now in the fortunate situation of having a huge amount of high-frequency financial data for a large number of assets. Over the past fifteen years the econometric methods to analyze the high-frequency data for a small number of assets has grown exponentially. At the same time the field of large dimensional data analysis has exploded providing us with a variety of tools to analyze a large cross-section of financial assets over a long time horizon. This paper merges these two literatures by developing statistical methods for estimating the systematic pattern in high frequency data for a large cross-section. One of the most popular methods for analyzing large cross-sectional data sets is factor analysis. Some of the most influential economic theories, e.g. the arbitrage pricing theory of Ross (1976) are based on factor models. While there is a well-developed inferential theory for factor models of large dimension with long time horizon and for factor models of small dimension based on high-frequency observations, the inferential theory for large dimensional high-frequency factor models is an area of active research.

This paper develops the statistical inferential theory for approximate factor models of large dimensions based on high-frequency observations. Conventional factor analysis requires a long time horizon, while this methodology also works with short time horizons, e.g. a week. If a large cross-section of firms and sufficiently many high-frequency asset prices are available, we can estimate the number of systematic factors and derive consistent and asymptotically mixed-normal estimators of the latent loadings and factors. These results are obtained for very general stochastic processes, namely Itô semimartingales with jumps, and an approximate factor structure which allows for weak serial and cross-sectional correlation in the idiosyncratic errors. The estimation approach can separate factors for systematic large sudden movements, so-called jumps factors, from continuous factors.

This methodology has many important applications as it can help us to understand systematic risk better. First, we obtain guidance on how many factors might explain the systematic movements and see how this number changes over short time horizons. Second, we can analyze how loadings and factors change over short time horizons and study their persistence. Third, we can analyze how continuous systematic risk factors, which capture the variation during "normal" times, are different from jump factors, which can explain systematic tail events. Fourth, after identifying the systematic and idiosyncratic components we can apply these two components separately to previous empirical high-frequency studies to see if there is a different effect for systematic versus nonsystematic movements. For example we can examine which components drive the leverage effect. In a separate paper, Pelger (2015), I apply my estimation method to a large high-frequency data set of the S&P500 firms to test these questions empirically.

My estimator for the loadings and factors is essentially the well-known principal component based estimator of Bai (2003), where I just use properly rescaled increments for the covariance estimation. However, except for very special cases the necessary assumptions and the proofs cannot be mapped into the long-horizon factor model and hence require new derivations. The asymptotic distribution results are in general different from the long-horizon factor model.[1] Furthermore conventional factor analysis does not distinguish between continuous and jump risk. Using a truncation approach, I can separate the continuous and jump components of the price processes, which I use to construct a "jump covariance" and a "continuous risk covariance" matrix. The latent continuous and jump factors can be separately estimated by principal component analysis.

This paper develops a new estimator for the number of factors that requires only the same weak assumptions as the loadings estimator in my model. The basic idea in most estimation approaches is that the systematic eigenvalues of the estimated covariance matrix or quadratic covariation matrix will explode, while the other eigenvalues of the idiosyncratic part will be bounded. Prominent estimators with good performance in simulations[2] impose the additional strong assumptions of random matrix theory that imply that a certain fraction of the small eigenvalues will be bounded from below and above and the largest residual eigenvalues will cluster. I propose the novel idea of perturbing the eigenvalues before analyzing the eigenvalue ratio. As long as the eigenvalue ratio of the perturbed eigenvalues is close to one, the spectrum is due to the residuals. Due to a weaker rate argument and not the strong assumptions of random matrix theory the eigenvalue ratio of perturbed idiosyncratic eigenvalues will cluster. The important contribution of my estimator is that it can estimate the number of continuous, jump and total factors separately and that it can deal with strong and weak factors as we are focussing on the residual spectrum. The approach is robust to the choice of the perturbation value. Simulations illustrate the excellent performance of my new estimator.

I extend my model into two directions. First, I include microstructure noise and develop an estimator for the variance of microstructure noise and for the impact of microstructure noise on the spectrum of the factor estimator, allowing us to test if a frequency is sufficiently coarse to neglect the noise. Second, I develop a new test to determine if a set of estimated statistical factors can be written as a linear combination of observed economic variables. The challenge is that factor models are only identified up to invertible transformations. I provide a new measure for the distance between two sets of factors and develop its asymptotic distribution under the

[1] (1) After rescaling the increments, we can interpret the quadratic covariation estimator as a sample covariance estimator. However, in contrast to the covariance estimator, the limiting object will be a random variable and the asymptotic distribution results have to be formulated in terms of stable convergence in law, which is stronger than convergence in distribution. (2) Models with jumps have "heavy-tailed rescaled increments" which cannot be accommodated in the relevant long-horizon factor models. (3) In stochastic volatility or stochastic intensity jump models the data is non-stationary. Some of the results in large dimensional factor analysis do not apply to non-stationary data. (4) In contrast to long-horizon factor analysis the asymptotic distribution of my estimators have a mixed Gaussian limit and so will generally have heavier tails than a normal distribution.

[2] E.g. Onatski (2010) and Ahn and Horenstein (2013)

same weak assumptions as for the estimation of the factors.

My work builds on the fast growing literatures in the two separate fields of large-dimensional factor analysis and high-frequency econometrics.[3] The notion of an "approximate factor model" was introduced by Chamberlain and Rothschild (1983), which allowed for a non-diagonal covariance matrix of the idiosyncratic component. They applied principal component analysis to the population covariance. Connor and Korajczyk (1988, 1993) study the use of principal component analysis in the case of an unknown covariance matrix, which has to be estimated. The general case of a static large dimensional factor model is treated in Bai (2003). He develops an inferential theory for factor models for a large cross-section and long time horizons based on a principal component analysis of the sample covariance matrix. His paper is the closest to mine from this literature. As pointed out before for general continuous-time processes we cannot map the high-frequency problem into the long horizon model. Forni, Hallin, Lippi and Reichlin (2000) introduced the dynamic principal component method. Fan, Liao and Mincheva (2013) study an approximate factor structure with sparsity. Some of the most relevant estimators for the number of factors in large-dimensional factor models based on long-horizons are the Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) estimators.[4] The last two estimators perform well in simulations, but their arguments which are based on random matrix theory seem not to be transferable to our high-frequency problem without imposing unrealistically strong assumptions on the processes.[5] Many of my asymptotic results for the estimation of the quadratic covariation are based on Jacod (2008), where he develops the asymptotic properties of realized power variations and related functionals of semimartingales. Aït-Sahalia and Jacod (2009a), Lee and Mykland (2008) and Mancini (2009) introduce a threshold estimator for separating the continuous from the jump variation, which I use in this paper.[6] Bollerslev and Todorov (2010) develop the theoretical framework for high-frequency factor models for a low dimension. Their results are applied empirically in Bollerslev, Li and Todorov (2015).

So far there are relatively few papers combing high-frequency analysis with high-dimensional regimes, but this is an active and growing literature. Important recent papers include Wang and Zou (2010), Tao, Wang and Chen (2013), and Tao, Wang and Zhou (2013) who establish results for large sparse matrices estimated with high-frequency observations. Fan, Furger and

[3]Bai and Ng (2008) provide a good overview of large dimensional factor analysis. An excellent and very up-to-date textbook treatment of high-frequency econometrics is Aït-Sahalia and Jacod (2014).

[4]There are many alternative methods, e.g. Hallin and Lisak (2007), Aumengual and Watson (2007), Alessi et al. (2010) or Kapetanious (2010), but in simulations they do not seem to outperform the above methods.

[5]The Bai and Ng (2002) paper uses an information criterion, while Onatski applies an eigenvalue difference estimator and Ahn and Horenstein an eigenvalue ratio approach. If the first systematic factors are stronger than other weak systematic factors the Ahn and Horenstein method can fail in simulations with realistic values, while the Onatski method can perform better as it focuses only on the residual eigenvalues.

[6]In an influential series of papers, Barndorff-Nielsen and Shephard (2004b, 2006) and Barndorff-Nielsen, Shephard, and Winkel (2006) introduce the concept of (bi-)power variation - a simple but effective technique to identify and measure the variation of jumps from intraday data.

Xiu (2014) estimate a large-dimensional covariance matrix with high-frequency data for a given factor structure. My results were derived simultaneously and independently to results in the two papers by Aït-Sahalia and Xiu (2015a+b). Their papers and my work both address the problem of finding structure in high-frequency financial data, but proceed in somewhat different directions and achieve complementary results. In their first paper Aït-Sahalia and Xiu (2015a) develop the inferential theory of principal component analysis applied to a low-dimensional cross-section of high-frequency data. I work in a large-dimensional setup which requires the additional structure of a factor model and derive the inferential theory for both the continuous and jump structures. In their second paper Aït-Sahalia and Xiu (2015b) considers a large-dimensional high-frequency factor model and they derive consistent estimators for the factors based on continuous processes.[7] Their identification is based on a sparsity assumption on the idiosyncratic covariance matrix. My main identification condition is a bounded eigenvalue condition on the idiosyncratic covariance matrix which allows me to also consider jumps and to derive the asymptotic distribution theory of the estimators.

The rest of the paper is organized as follows. Section 2 introduces the factor model. In Section 3 I explain my estimators. Section 4 summarizes the assumptions and the asymptotic consistency results for the estimators of the factors, loadings and common components. In Subsection 4.3 I also deal with the separation into continuous and jump factors. In Section 5 I show the asymptotic mixed-normal distribution of the estimators and derive consistent estimators for the covariance matrices occurring in the limiting distributions. In Section 6 I develop the estimator for the number of factors. The extension to microstructure noise is treated in Section 7. The test for comparing two sets of factors is presented in Section 8. Section 9 discusses the differences with long horizon models and Section 10 presents some simulation results. Concluding remarks are provided in Section 11. All the proofs are deferred to the appendices.

2 Model Setup

Assume the N-dimensional stochastic process $X(t)$ can be explained by a factor model, i.e.

$$X_i(t) = \Lambda_i^\top F(t) + e_i(t) \qquad i = 1, ..., N \text{ and } t \in [0, T]$$

where Λ_i is a $K \times 1$ dimensional vector and $F(t)$ is a K-dimensional stochastic process in continuous time. The loadings Λ_i describe the exposure to the systematic factors F, while the residuals e_i are stochastic processes that describe the idiosyncratic component. $X(t)$ will

[7] Aït-Sahalia and Xiu (2015b) also develop an estimator for the number of factors which is essentially an extension of the Bai and Ng (2002) estimator to high-frequency data. Aït-Sahalia and Xiu's techniques assume continuous processes. I also allow for jumps and my approach can deal with strong and weak factors.

typically be the log-price process. However, we only observe the stochastic process X at M discrete time observations in the interval $[0,T]$. If we use an equidistant grid[8], we can define the time increments as $\Delta_M = t_{j+1} - t_j = \frac{T}{M}$ and observe

$$X_i(t_j) = \Lambda_i^\top F(t_j) + e_i(t_j) \qquad i = 1, ..., N \text{ and } j = 1, ..., M$$

or in vector notation

$$X(t_j) = \Lambda F(t_j) + e(t_j) \qquad j = 1, ..., M.$$

with $\Lambda = (\Lambda_1, ..., \Lambda_N)^\top$. In my setup the number of cross-sectional observations N and the number of high-frequency observations M is large, while the time horizon T and the number of systematic factors K is fixed. The loadings Λ, factors F, residuals e and number of factors K are unknown and have to be estimated.

3 Estimation Approach

We have M observations of the N-dimensional stochastic process X in the time interval $[0,T]$. For the time increments $\Delta_M = \frac{T}{M} = t_{j+1} - t_j$ we denote the increments of the stochastic processes by

$$X_{j,i} = X_i(t_{j+1}) - X_i(t_j) \qquad F_j = F(t_{j+1}) - F(t_j) \qquad e_{j,i} = e_i(t_{j+1}) - e_i(t_j).$$

In matrix notation we have

$$\underset{(M\times N)}{X} = \underset{(M\times K)}{F}\underset{(K\times N)}{\Lambda^\top} + \underset{(M\times N)}{e}.$$

For a given K our goal is to estimate Λ and F. As in any factor model where only X is observed Λ and F are only identified up to K^2 parameters as $F\Lambda^\top = FAA^{-1}\Lambda^\top$ for any arbitrary invertible $K \times K$ matrix A. Hence, for my estimator I impose the K^2 standard restrictions that $\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} = I_K$ which gives us $\frac{K(K+1)}{2}$ restrictions and that $\hat{F}^\top\hat{F}$ is a diagonal matrix, which yields another $\frac{K(K-1)}{2}$ restrictions.

Denote the K largest eigenvalues of $\frac{1}{N}X^\top X$ by V_{MN}. The estimator for the loadings $\hat{\Lambda}$ is defined as the K eigenvectors of V_{MN} multiplied by $\sqrt{N}$. The estimator for the factor increments is $\hat{F} = \frac{1}{N}X\hat{\Lambda}$. Note that $\frac{1}{N}X^\top X$ is an estimator for $\frac{1}{N}[X,X]$ for a finite N. We study the asymptotic theory for $M, N \to \infty$. As in Bai (2003) we consider a simultaneous limit which allows (N,M) to increase along all possible paths.

[8] Most of my results would go through under a time grid that is not equidistant as long as the largest time increment goes to zero with speed $O\left(\frac{1}{M}\right)$.

The systematic component of $X(t)$ is the part that is explained by the factors and defined as $C(t) = \Lambda F(t)$. The increments of the systematic component $C_{j,i} = F_j \Lambda_i^\top$ are estimated by $\hat{C}_{j,i} = \hat{F}_j \hat{\Lambda}_i^\top$.

We are also interested in estimating the continuous component, jump component and the volatility of the factors. Denoting by F^C the factors that have a continuous component and by F^D the factor processes that have a jump component, we can write

$$X(t) = \Lambda^C F^C(t) + \Lambda^D F^D(t) + e(t).$$

Note, that for factors that have both, a continuous and a jump component, the corresponding loadings have to coincide. In the following we assume a non-redundant representation of the K^C continuous and K^D jump factors. For example if we have K factors which have all exactly the same jump component but different continuous components, this results in K different total factors and $K^C = K$ different continuous factors, but in only $K^D = 1$ jump factor.

Intuitively under some assumptions we can identify the jumps of the process $X_i(t)$ as the big movements that are larger than a specific threshold. Set the threshold identifier for jumps as $\alpha \Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}_{j,i}^C = X_{j,i} \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}$ and $\hat{X}_{j,i}^D = X_{j,i} \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta_M^{\bar{\omega}}\}}$.[9] The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X.

The quadratic covariation of the factors can be estimated by $\hat{F}^\top \hat{F}$ and the volatility component of the factors by $\hat{F}^{C\top} \hat{F}^C$. I show that the estimated increments of the factors $\hat{F}, \hat{F}^C$ and $\hat{F}^D$ can be used to estimate the quadratic covariation with any other process.

The number of factors can be consistently estimated through the perturbed eigenvalue ratio statistic and hence, we can replace the unknown number K by its estimator $\hat{K}$. Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. We choose a slowly increasing sequence $g(N, M)$ such that $\frac{g(N,M)}{N} \to 0$ and $g(N, M) \to \infty$. Based on simulations a good choice for the perturbation term g is the median eigenvalue rescaled by $\sqrt{N}$. Then, we define perturbed eigenvalues $\hat{\lambda}_k = \lambda_k + g(N, M)$ and the perturbed eigenvalue ratio statistic

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \text{for } k = 1, ..., N-1.$$

The estimator for the number of factors is defined as the first time that the perturbed eigenvalue ratio statistic does not cluster around 1 any more:

$$\hat{K}(\gamma) = \max\{k \leq N-1 : ER_k > 1 + \gamma\} \qquad \text{for } \gamma > 0.$$

[9]Choices of α and $\bar{\omega}$ are standard in the literature (see, e.g. Aït-Sahalia and Jacod (2014)) and are discussed below when implemented in simulations.

If $ER_k < 1+\gamma$ for all k, then we set $\hat{K}(c) = 0$. The definition of $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ is analogous but using λ_i^C respectively λ_i^D of the matrices $\hat{X}^{C\top}\hat{X}^C$ and $\hat{X}^{D\top}\hat{X}^D$. Based on extensive simulations a constant γ between 0.05 and 0.2 seems to be good choice.

4 Consistency Results

4.1 Assumptions on Stochastic Processes

All the stochastic processes considered in this paper are locally bounded special Itô semimartingales as defined in Definition 1 in Appendix B. These particular semimartingales are the most general stochastic processes for which we can develop an asymptotic theory for the estimator of the quadratic covariation. A d-dimensional locally bounded special Itô semimartingale Y can be represented as

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \delta(s,x)(\mu - \nu)(ds, dx)$$

where b_s is a locally bounded predictable drift term, σ_s is an adapted cádlág volatility process, W is a d-dimensional Brownian motion and $\int_0^t \int_E \delta(s,x)(\mu - \nu)(ds, dx)$ describes a jump martingale. μ is a Poisson random measure on $\mathbb{R}_+ \times E$ with $(E, \mathbb{E})$ an auxiliary measurable space on the space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$. The predictable compensator (or intensity measure) of μ is $\nu(ds, dx) = ds \times v(dx)$ for some given finite or sigma-finite measure on $(E, \mathbb{E})$. These dynamics are very general and completely non-parametric. They allow for correlation between the volatility and asset price processes. I only impose some week regularity conditions in Definition 1. The model includes many well-known continuous-time models as special cases: for example stochastic volatility models like the CIR or Heston model, the affine class of models in Duffie, Pan and Singleton (2000), Barndorff-Nielsen and Shephard's (2002) Ornstein-Uhlenbeck stochastic volatility model with jumps or Andersen, Benzoni, and Lund's (2002) stochastic volatility model with log-normal jumps generated by a non-homogenous Poisson process.

I denote by $\Delta_j Y$ the jth observed increment of the process Y, i.e. $\Delta_j Y = Y(t_{j+1}) - Y(t_j)$ and write $\Delta Y(t) = Y(t) - Y(t-)$ for the jumps of the process Y. Of course, $\Delta Y(t) = 0$ for all $t \in [0, T]$ if the process is continuous. The sum of squared increments converges to the quadratic covariation for $M \to \infty$:

$$\sum_{j=1}^{M} \Delta_j Y_i \Delta_j Y_k \xrightarrow{p} [Y_i, Y_k]. \qquad i, k = 1, ..., d.$$

The predictable quadratic covariation $\langle Y_i, Y_k \rangle$ is the predictable conditional expectation of $[Y_i, Y_k]$, i.e. it is the so-called compensator process. It is the same as the realized quadratic covariation $[X_i, X_k]$ for a continuous process, but differs if the processes have jumps. The realized quadratic covariation $[Y_i, Y_k]_t$ and the conditional quadratic covariation $\langle Y_i, Y_k \rangle_t$ are themselves stochastic processes. If I leave out the time index t, it means that I am considering the quadratic covariation evaluated at the terminal time T, which is a random variable. For more details see Rogers (2004) or Jacod and Shiryaev (2002).

4.2 Consistency

The key assumption for obtaining a consistent estimator for the loadings and factors is an approximate factor structure. It requires that the factors are systematic in the sense that they cannot be diversified away, while the idiosyncratic residuals are nonsystematic and can be diversified away. The approximate factor structure assumption uses the idea of appropriately bounded eigenvalues of the residual quadratic covariation matrix, which is analogous to Chamberlain and Rothschild (1983) and Chamberlain (1988). Let $\|A\| = (tr(A \top A))^{1/2}$ denote the norm of a matrix A and $\lambda_i(A)$ the i's largest singular value of the matrix A, i.e. the square-root of the i's largest eigenvalue of $A^\top A$. If A is a symmetric matrix then λ_i is simply the i's largest eigenvalue of A.

Assumption 1. Factor structure assumptions

1. **Underlying stochastic processes**
   F and e_i are Itô-semimartingales as defined in Definition 1

$$F(t) = F(0) + \int_0^t b_F(s)ds + \int_0^t \sigma_F(s)dW_s + \sum_{s \le t} \Delta F(s)$$
$$e_i(t) = e(0) + \int_0^t b_{e_i}(s)ds + \int_0^t \sigma_{e_i}(s)dW_s + \sum_{s \le t} \Delta e_i(s)$$

   In addition each e_i is a square integrable martingale.

2. **Factors and factor loadings**
   The quadratic covariation matrix of the factors Σ_F is positive definite a.s.

$$\sum_{j=1}^{M} F_j F_j^\top \xrightarrow{p} [F, F]_T =: \Sigma_F$$

   and

$$\left\| \frac{\Lambda^\top \Lambda}{N} - \Sigma_\Lambda \right\| \to 0.$$

where the matrix Σ_Λ is also positive definite. The loadings are bounded, i.e. $\|\Lambda_i\| < \infty$ for all $i = 1, ..., N$.

3. **Independence of F and e**

   The factor process F and the residual processes e are independent.

4. **Approximate factor structure**

   The largest eigenvalue of the residual quadratic covariation matrix is bounded in probability, i.e.

$$\lambda_1([e,e]) = O_p(1).$$

   As the predictable quadratic covariation is absolutely continuous, we can define the instantaneous predictable quadratic covariation as

$$\frac{d\langle e_i, e_k\rangle_t}{dt} = \sigma_{e_{i,k}}(t) + \int \delta_{e_{i,k}}(z) v_t(z) =: G_{i,k}(t).$$

   We assume that the largest eigenvalue of the matrix $G(t)$ is almost surely bounded for all t:

$$\lambda_1(G(t)) < C \qquad \textit{a.s. for all } t \textit{ for some constant } C.$$

5. **Identification condition** *All Eigenvalues of $\Sigma_\Lambda \Sigma_F$ are distinct a.s..*

The most important part of Assumption 1 is the approximate factor structure in point 4. It implies that the residual risk can be diversified away. Point 1 states that we can use the very general class of stochastic processes defined in Definition 1. The assumption that the residuals are martingales and hence do not have a drift term is only necessary for the asymptotic distribution results. The consistency results do not require this assumption. Point 2 implies that the factors affect an infinite number of assets and hence cannot be diversified away. Point 3 can be relaxed to allow for a weak correlation between the factors and residuals. This assumption is only used to derive the asymptotic distribution of the estimators. The approximate factor structure assumption in point 4 puts a restriction on the correlation of the residual terms. It allows for cross-sectional (and also serial) correlation in the residual terms as long as it is not too strong. We can relax the approximate factor structure assumption. Instead of almost sure boundedness of the predictable instantaneous quadratic covariation matrix of the residuals it is sufficient to assume that

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k\neq i}^{N} \Lambda_i G_{i,k}(t) \Lambda_k^\top < C \qquad \text{a.s. for all } t$$

Then, all main results except for Theorem 6 and 9 continue to hold. Under this weaker assumption we do not assume that the diagonal elements of G are almost surely bounded. By Definition 1 the diagonal elements of G are already locally bounded which is sufficient for most of our results.

Note that point 4 puts restrictions on both the realized and the conditional quadratic covariation matrix. In the case of continuous residual processes, the conditions on the conditional quadratic covariation matrix are obviously sufficient. However, in our more general setup it is not sufficient to restrict only the conditional quadratic covariation matrix.

Assumption 2. Weak dependence of error terms
The row sum of the quadratic covariation of the residuals is bounded in probability:

$$\sum_{i=1}^{N} \|[e_k, e_i]\| = O_p(1) \qquad \forall k = 1, ..., N$$

Assumption 2 is stronger than $\lambda_1([e,e]) = O_p(1)$ in Assumption 1. As the largest eigenvector of a matrix can be bounded by the largest absolute row sum, Assumption 2 implies $\lambda_1([e,e]) = O_p(1)$. If the residuals are cross-sectionally independent it is trivially satisfied. However it allows for a weak correlation between the residual processes. For example, if the residual part of each asset is only correlated with a finite number of residuals of other assets, it will be satisfied.

As pointed out before, the factors F and loadings Λ are not separately identifiable. However, we can estimate them up to an invertible $K \times K$ matrix H. Hence, my estimator $\hat{\Lambda}$ will estimate ΛH and $\hat{F}$ will estimate $F H^{\top^{-1}}$. Note, that the common component is well-identified and $\hat{F}\hat{\Lambda}^\top = \hat{F} H^{\top^{-1}} H^\top \Lambda^\top$. For almost all purposes knowing ΛH or $F H^{\top^{-1}}$ is as good as knowing Λ or F as what is usually of interest is the vector space spanned by the factors. For example testing the significance of F or $F H^{\top^{-1}}$ in a linear regression yields the same results.[10]

In my general approximate factor models we require N and M to go to infinity. The rates of convergence will usually depend on the smaller of these two values denoted by $\delta = \min(N, M)$. As noted before we consider a simultaneous limit for N and M and not a path-wise or sequential limit. Without further assumptions the asymptotic results do not hold for a fixed N or M. In this sense the large dimension of our problem, which makes the analysis more complicated, also helps us to obtain more general results and turns the “curse of dimensionality” into a “blessing”.

Note that F_j is the increment $\Delta_j F$ and goes to zero for $M \to \infty$ for almost all increments. It can be shown that in a specific sense we can also consistently estimate the factor increments, but the asymptotic statements will be formulated in terms of the stochastic process F evaluated

[10] For a more detailed discussion see Bai (2003) and Bai and Ng (2008).

at a discrete time point t_j. For example $F_T = \sum_{j=1}^{M} F_j$ denotes the factor process evaluated at time T. Similarly we can evaluate the process at any other discrete time point $T_m = m \cdot \Delta_M$ as long as $m \cdot \Delta_M$ does not go to zero. Essentially m has to be proportional to M. For example, we could chose T_m equal to $\frac{1}{2}T$ or $\frac{1}{4}T$. The terminal time T can always be replaced by the time T_m in all the theorems. The same holds for the common component.

Theorem 1. Consistency of estimators:
Define the rate $\delta = \min(N, M)$ *and the invertible matrix* $H = \frac{1}{N}\left(F^\top F\right)\left(\Lambda^\top \hat{\Lambda}\right) V_{MN}^{-1}$. *Then the following consistency results hold:*

1. *Consistency of loadings estimator: Under Assumption 1 it follows that*

$$\hat{\Lambda}_i - H^\top \Lambda_i = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

2. *Consistency of factor estimator and common component: Under Assumptions 1 and 2 it follows that*

$$\hat{F}_T - H^{-1} F_T = O_p\left(\frac{1}{\sqrt{\delta}}\right), \qquad \hat{C}_{T,i} - C_{T,i} = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

3. *Consistency of quadratic variation: Under Assumptions 1 and 2 and for any stochastic process* $Y(t)$ *satisfying Definition 1 we have for* $\frac{\sqrt{M}}{N} \to 0$ *and* $\delta \to \infty$*:*

$$\sum_{j=1}^{M} \hat{F}_j \hat{F}_j^\top = H^{-1}[F,F]_T H^{-1^\top} + o_p(1), \qquad \sum_{j=1}^{M} \hat{F}_j Y_j = H^{-1}[F,Y]_T + o_p(1)$$

$$\sum_{j=1}^{M} \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1), \qquad \sum_{j=1}^{M} \hat{e}_{j,i} Y_j = [e_i, Y] + o_p(1)$$

$$\sum_{j=1}^{M} \hat{C}_{j,i}\hat{C}_{j,k} = [C_i, C_k] + o_p(1), \qquad \sum_{j=1}^{M} \hat{C}_{j,i} Y_j = [C_i, Y] + o_p(1).$$

for $i, k = 1, ..., N$.

This statement only provides a pointwise convergence of processes evaluated at specific times. A stronger statement would be to show weak convergence for the stochastic processes. However, weak convergence of stochastic processes requires significantly stronger assumptions[11] and will in general not be satisfied under my assumptions.

[11] See for example Prigent (2003)

4.3 Separating Continuous and Jump Factors

Using a thresholding approach we can separate the continuous and jump movements in the observable process X and estimate the systematic continuous and jump factors. The idea is that with sufficiently many high-frequency observations, we can identify the jumps in X as the movements that are above a certain threshold. This allows us to separate the quadratic covariation matrix of X into its continuous and jump component. Then applying principal component analysis to each of these two matrices we obtain our separate factors. A crucial assumption is that the thresholding approach can actually identify the jumps:

Assumption 3. Truncation identification
F and e_i have only finite activity jumps and factor jumps are not "hidden" by idiosyncratic jumps:

$$\mathbb{P}\left(\Delta X_i(t) = 0 \text{ if } \Delta(\Lambda_i^\top F(t)) \neq 0 \text{ and } \Delta e_i(t) \neq 0\right) = 0.$$

The quadratic covariation matrix of the continuous factors $[F^C, F^C]$ and of the jump factors $[F^D, F^D]$ are each positive definite a.s. and the matrices $\frac{{\Lambda^C}^\top \Lambda^C}{N}$ and $\frac{{\Lambda^D}^\top \Lambda^D}{N}$ each converge in probability to positive definite matrices.

Assumption 3 has three important parts. First, we require the processes to have only finite jump activity. This mean that on every finite time interval there are almost surely only finitely many jumps. With infinite activity jump processes, i.e. each interval can contain infinitely many small jumps, we cannot separate the continuous and discontinuous part of a process. Second, we assume that a jump in the factors or the idiosyncratic part implies a jump in the process X_i. The reverse is trivially satisfied. This second assumption is important to identify all times of discontinuities of the unobserved factors and residuals. This second part is always satisfied as soon as the Lévy measure of F_i and e_i have a density, which holds in most models used in the literature. The third statement is a non-redundancy condition and requires each systematic jump factor to jump at least once in the data. This is a straightforward and necessary condition to identify any jump factor. Hence, the main restriction in Assumption 3 is the finite jump activity. For example compound poisson processes with stochastic intensity rate fall into this category.

Theorem 2. Separating continuous and jump factors:
Assume Assumptions 1 and 3 hold. Set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}_{j,i}^C = X_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}_{j,i}^D = X_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}\}}$. The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X. Define $H^C = \frac{1}{N}\left({F^C}^\top F^C\right)\left({\Lambda^C}^\top \hat{\Lambda}^C\right){V_{MN}^C}^{-1}$ and $H^D = \frac{1}{N}\left({F^D}^\top F^D\right)\left({\Lambda^D}^\top \hat{\Lambda}^D\right){V_{MN}^D}^{-1}$.

1. *The continuous and jump loadings can be estimated consistently:*

$$\hat{\Lambda}_i^C = {H^C}^\top \Lambda_i^C + o_p(1) \qquad , \qquad \hat{\Lambda}_i^D = {H^D}^\top \Lambda_i^D + o_p(1).$$

2. *Assume that additionally Assumption 2 holds. The continuous and jump factors can only be estimated up to a finite variation bias term*

$$\hat{F}_T^C = {H^C}^{-1} F_T^C + o_p(1) + \textit{finite variation term}$$
$$\hat{F}_T^D = {H^D}^{-1} F_T^D + o_p(1) + \textit{finite variation term.}$$

3. *Under the additional Assumption 2 we can estimate consistently the covariation of the continuous and jump factors with other processes. Let $Y(t)$ be an Itô-semimartingale satisfying Definition 1. Then we have for $\frac{\sqrt{M}}{N} \to 0$ and $\delta \to \infty$:*

$$\sum_{j=1}^M \hat{F}_j^C Y_j = {H^C}^{-1} [F^C, Y]_T + o_p(1) \qquad , \qquad \sum_{j=1}^M \hat{F}_j^D Y_j = {H^D}^{-1} [F^D, Y]_T + o_p(1).$$

The theorem states that we can estimate the factors only up to a finite variation term, i.e. we can only estimate the martingale part of the process correctly. The intuition behind this problem is simple. The truncation estimator can correctly separate the jumps from the continuous martingale part. However, all the drift terms will be assigned to the continuous component. If a jump factor also has a drift term, this will now appear in the continuous part and as this drift term affects infinitely many cross-sectional X_i, it cannot be diversified away.

5 Asymptotic Distribution

5.1 Distribution Results

The assumptions for asymptotic mixed-normality of the estimators are stronger than those needed for consistency. Although asymptotic mixed-normality of the loadings does not require additional assumptions, the asymptotic normality of the factors needs substantially stronger assumptions. This should not be surprising as essentially all central limit theorems impose restrictions on the tail behavior of the sampled random variables.

In order to obtain a mixed Gaussian limit distribution for the loadings we need to assume that there are no common jumps in σ_F and e_i and in σ_{e_i} and F. Without this assumption the estimator for the loadings still converges at the same rate, but it is not mixed-normally distributed any more. Note that Assumption 1 requires the independence of F and e, which implies the no common jump assumption.

Theorem 3. Asymptotic distribution of loadings
Assume Assumptions 1 and 2 hold and define $\delta = min(N, M)$. *Then*

$$\sqrt{M}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = V_{MN}^{-1}\left(\frac{\hat{\Lambda}^\top \Lambda}{N}\right)\sqrt{M}F^\top e_i + O_p\left(\frac{\sqrt{M}}{\delta}\right)$$

1. *If* $\frac{\sqrt{M}}{N} \to 0$, *then*

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N\left(0, V^{-1}Q\Gamma_i Q^\top V^{-1}\right)$$

   where V *is the diagonal matrix of eigenvalues of* $\Sigma_\Lambda^{\frac{1}{2}}\Sigma_F\Sigma_\Lambda^{\frac{1}{2}}$ *and* $\underset{N,M\to\infty}{plim}\frac{\hat{\Lambda}^\top\Lambda}{N} = Q = V^{\frac{1}{2}}\Upsilon^\top\sigma_F^{\frac{1}{2}}$ *with* Υ *being the eigenvectors of* V. *The entry* $\{l, g\}$ *of the* $K \times K$ *matrix* Γ_i *is given by*

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g}\sigma_{e_i}^2 ds + \sum_{s\leq T}\Delta F^l(s)\Delta F^g(s)\sigma_{e_i}^2(s) + \sum_{s'\leq T}\Delta e_i^2(s')\sigma_{F^g,F^l}(s').$$

   F^l *denotes the* l*-th component of the the* K *dimensional process* F *and* σ_{F^l,F^g} *are the entries of its* $K \times K$ *dimensional volatility matrix.*

2. *If* $\liminf \frac{\sqrt{M}}{N} \geq \tau > 0$, *then* $N(\hat{\Lambda}_i - \Lambda_i H) = O_p(1)$.

The asymptotic expansion is very similar to the conventional factor analysis in Bai (2003), but the limiting distributions of the loadings is obviously different. The mode of convergence is stable convergence in law, which is stronger than simple convergence in distribution.[12] Here we can see very clearly how the results from high-frequency econometrics impact the estimators in our factor model.

Assumption 4. Asymptotically negligible jumps of error terms
Assume Z *is some continuous square integrable martingale with quadratic variation* $\langle Z, Z\rangle_t$. *Assume that the jumps of the martingale* $\frac{1}{\sqrt{N}}\sum_{i=1}^N e_i(t)$ *are asymptotically negligible in the sense that*

$$\frac{\Lambda^\top[e,e]_t\Lambda}{N} \xrightarrow{p} \langle Z, Z\rangle_t\,, \qquad \frac{\Lambda^\top\langle e^D, e^D\rangle_t\Lambda}{N} \xrightarrow{p} 0 \qquad \forall t > 0.$$

Assumption 4 is needed to obtain an asymptotic mixed-normal distribution for the factor estimator. It means that only finitely many residual terms can have a jump component. Hence, the weighted average of residual terms has a quadratic covariation that depends only on the continuous quadratic covariation. This assumption is essentially a Lindeberg condition. If

[12] For more details see Aït-Sahalia and Jacod (2014).

it is not satisfied and under additional assumptions the factor estimator converges with the same rate to a distribution with the same variance, but with heavier tails than a mixed-normal distribution.

Assumption 5. Weaker dependence of error terms

- **Assumption 5.1: Weak serial dependence**
  The error terms exhibit weak serial dependence if and only if

$$\left\| \mathbb{E}\left[e_{ji} e_{jr} \sum_{l \neq j} e_{li} \sum_{s \neq j} e_{sr} \right] \right\| \leq C \left\| \mathbb{E}[e_{ji} e_{jr}] \right\| \left\| \mathbb{E}\left[\sum_{l \neq j} e_{li} \sum_{s \neq j} e_{lr} \right] \right\|$$

  for some finite constant C and for all $i, r = 1, ..., N$ and for all partitions $[t_1, ..., t_M]$ of $[0, T]$.
- **Assumption 5.2: Weak cross-sectional dependence**
  The error terms exhibit weak cross-sectional dependence if and only if

$$\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[e_{ji}^2 e_{jr}^2 \right] = O\left(\frac{1}{\delta} \right)$$

  for all $i, r = 1, ..., N$ and for all partitions $[t_1, ..., t_M]$ of $[0, T]$ for $M, N \to \infty$ and

$$\sum_{i=1}^{N} |G_{k,i}(t)| \leq C \qquad a.s. \text{ for all } k = 1, ..., N \text{ and } t \in (0, T] \text{ and some constant } C.$$

Assumption 5 is only needed to obtain the general rate results for the asymptotic distribution of the factors. If $\frac{N}{M} \to 0$, we don't need it anymore. Lemma 1 gives sufficient conditions for this assumption. Essentially, if the residual terms are independent and "almost" continuous then it holds. Assumption 5 is not required for any consistency results.

Lemma 1. Sufficient conditions for weaker dependence
Assume Assumptions 1 and 2 hold and that

1. *e_i has independent increments.*
2. *e_i has 4th moments.*
3. *$\mathbb{E}\left[\sum_{i=1}^{N} \langle e_i^D, e_i^D \rangle \right] \leq C$ for some constant C and for all N.*
4. *$\sum_{i=1}^{N} |G_{k,i}(t)| \leq C$ a.s. for all $k = 1, ..., N$ and $t \in (0, T]$ and some constant C.*

Then Assumption 5 is satisfied.

Theorem 4. Asymptotic distribution of the factors:
Assume Assumptions 1 and 2 hold. Then

$$\sqrt{N}\left(\hat{F}_T - H^{-1}F_T\right) = \frac{1}{\sqrt{N}} e_T \Lambda H + O_P\left(\frac{\sqrt{N}}{\sqrt{M}}\right) + O_p\left(\frac{\sqrt{N}}{\delta}\right)$$

If Assumptions 4 and 5 hold and $\frac{\sqrt{N}}{M} \to 0$ *or only Assumption 4 holds and* $\frac{N}{M} \to 0$*:*

$$\sqrt{N}\left(\hat{F}_T - H^{-1}F_T\right) \xrightarrow{L-s} N\left(0, {Q^{-1}}^\top \Phi_T Q^{-1}\right)$$

with $\Phi_T = \underset{N\to\infty}{plim} \frac{\Lambda^\top [e] \Lambda}{N}$.

The assumptions needed for Theorem 4 are stronger than for all the other theorems. Although they might not always be satisfied in practice, simulations indicate that the asymptotic distribution results still seem to provide a very good approximation even if the conditions are violated. As noted before it is possible to show that under weaker assumptions the factor estimators have the same rate and variance, but an asymptotic distribution that is different from a mixed-normal distribution.

The next theorem about the common components essentially combines the previous two theorems.

Theorem 5. Asymptotic distribution of the common components
Define $C_{T,i} = \Lambda_i^\top F_T$ *and* $\hat{C}_{T,i} = \hat{\Lambda}_i^\top \hat{F}_T$ *Assume that Assumptions 1 - 4 hold.*

1. *If Assumption 5 holds, i.e. weak serial dependence and cross-sectional dependence, then for any sequence* N, M

$$\frac{\sqrt{\delta}\left(\hat{C}_{T,i} - C_{T,i}\right)}{\sqrt{\frac{\delta}{N} W_{T,i} + \frac{\delta}{M} V_{T,i}}} \xrightarrow{D} N(0,1)$$

2. *Assume* $\frac{N}{M} \to 0$ *(but we do not require Assumption 5)*

$$\frac{\sqrt{N}\left(C_{T,i} - \hat{C}_{T,i}\right)}{\sqrt{W_{T,i}}} \xrightarrow{D} N(0,1)$$

with

$$W_{T,i} = \Lambda_i^\top \Sigma_\Lambda^{-1} \Phi_T \Sigma_\Lambda^{-1} \Lambda_i$$
$$V_{T,i} = F_T^\top \Sigma_F^{-1} \Gamma_i \Sigma_F^{-1} F_T.$$

5.2 Estimating Covariance Matrices

The asymptotic covariance matrix for the estimator of the loadings can be estimated consistently under relatively weak assumptions, while the asymptotic covariance of the factor estimator requires stricter conditions. In order to estimate the asymptotic covariance for the loadings, we cannot simply apply the truncation approach to the estimated processes. The asymptotic covariance matrix of the factors runs into a dimensionality problem, which can only be solved under additional assumptions.

Theorem 6. Feasible estimator of covariance matrix of loadings
Assume Assumptions 1 and 2 hold and $\frac{\sqrt{M}}{N} \to 0$. *Define the asymptotic covariance matrix of the loadings as* $\Theta_{\Lambda,i} = V^{-1} Q \Gamma_i Q^\top V^{-1}$. *Take any sequence of integers* $k \to \infty$, $\frac{k}{M} \to 0$. *Denote by* $I(j)$ *a local window of length* $\frac{2k}{M}$ *around* j. *Define the* $K \times K$ *matrix* $\hat{\Gamma}_i$ *by*

$$\begin{aligned}
\hat{\Gamma}_i =& M \sum_{j=1}^{M} \left(\frac{\hat{X}_j^C \hat{\Lambda}}{N} \right) \left(\frac{\hat{X}_j^C \hat{\Lambda}}{N} \right)^\top \left(\hat{X}_{j,i}^C - \frac{\hat{X}_j^C \hat{\Lambda}}{N} \hat{\Lambda}_i \right)^2 \\
&+ \frac{M}{2k} \sum_{j=k+1}^{M-k} \left(\frac{\hat{X}_j^D \hat{\Lambda}}{N} \right) \left(\frac{\hat{X}_j^D \hat{\Lambda}}{N} \right)^\top \left(\sum_{h \in I(j)} \left(\hat{X}_{h,i}^C - \frac{\hat{X}_h^C \hat{\Lambda}}{N} \hat{\Lambda}_i \right)^2 \right) \\
&+ \frac{M}{2k} \sum_{j=k+1}^{M-k} \left(\hat{X}_{j,i}^D - \frac{\hat{X}_j^D \hat{\Lambda}}{N} \hat{\Lambda}_i \right)^2 \left(\sum_{h \in I(j)} \left(\frac{\hat{X}_h^C \hat{\Lambda}}{N} \right) \left(\frac{\hat{X}_h^C \hat{\Lambda}}{N} \right)^\top \right)
\end{aligned}$$

Then a feasible estimator for $\Theta_{\Lambda,i}$ *is* $\hat{\Theta}_{\Lambda,i} = V_{MN}^{-1} \hat{\Gamma}_i V_{MN}^{-1} \xrightarrow{p} \Theta_{\Lambda,i}$ *and*

$$\sqrt{M} \hat{\Theta}_{\Lambda,i}^{-1/2} (\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{D} N(0, I_K).$$

Theorem 7. Consistent estimator of covariance matrix of factors
Assume the Assumptions of Theorem 4 hold and $\sqrt{N}\left(\hat{F}_T - H^{-1} F_T\right) \xrightarrow{L-s} N(0, \Theta_F)$ *with* $\Theta_F = \underset{N,M\to\infty}{plim} H^\top \frac{\Lambda^\top [e] \Lambda}{N} H$. *Assume that the error terms are cross-sectionally independent. Denote the estimator of the residuals by* $\hat{e}_{j,i} = X_{j,i} - \hat{C}_{j,i}$. *Then a consistent estimator is* $\hat{\Theta}_F = \frac{1}{N} \sum_{i=1}^N \hat{\Lambda}_i \hat{e}_i^\top \hat{e}_i \hat{\Lambda}_i^\top \xrightarrow{p} \Theta_F$ *and*

$$\sqrt{N} \hat{\Theta}_F^{-1/2} (\hat{F}_T - H^{-1} F_T) \xrightarrow{D} N(0, I_K).$$

The assumption of cross-sectional independence here is somewhat at odds with our general approximate factor model. The idea behind the approximate factor model is exactly to allow for weak dependence in the residuals. However, without further assumptions the quadratic covariation matrix of the residuals cannot be estimated consistently as its dimension is growing with N. Even if we knew the true residual process $e(t)$ we would still run into the same problem. Assuming cross-sectional independence is the simplest way to reduce the number of parameters that have to be estimated. We could extend this theorem to allow for a parametric model capturing the weak dependence between the residuals or we could impose a sparsity assumption similar to Fan, Liao and Mincheva (2013). In both cases the theorem would continue to hold.

Theorem 8. Consistent estimator of covariance matrix of common components
Assume Assumptions 1-5 hold and that the residual terms e are cross-sectionally independent. Then for any sequence N, M

$$\left(\frac{1}{N}\hat{W}_{T,i} + \frac{1}{M}\hat{V}_{T,i}\right)^{-1/2}\left(\hat{C}_{T,i} - C_{T,i}\right) \xrightarrow{D} N(0,1)$$

with $\hat{W}_{T,i} = \hat{\Lambda}_i^\top \hat{\Theta}_F \hat{\Lambda}_i$ *and* $\hat{V}_{T,i} = \hat{F}_T^\top \left(\hat{F}^\top \hat{F}\right)^{-1} \hat{\Gamma}_i \left(\hat{F}^\top \hat{F}\right)^{-1} \hat{F}_T$.

6 Estimating the Number of Factors

I have developed a consistent estimator for the number of total, continuous and jump factors, that does not require stronger assumptions than those needed for consistency. Intuitively the large eigenvalues are associated with the systematic factors and hence the problem of estimating the number of factors is roughly equivalent to deciding which eigenvalues are considered to be large with respect to the rest of the spectrum. Under the assumptions that we need for consistency I can show that the first K "systematic" eigenvalues of $X^\top X$ are $O_p(N)$, while the nonsystematic eigenvalues are $O_p(1)$. A straightforward estimator for the number of factors considers the eigenvalue ratio of two successive eigenvalues and associates the number of factors with a large eigenvalue ratio. However, without very strong assumptions we cannot bound the small eigenvalues from below, which could lead to exploding eigenvalue ratios in the nonsystematic spectrum. I propose a perturbation method to avoid this problem. As long as the eigenvalue ratios of the perturbed eigenvalues cluster, we are in the nonsystematic spectrum. As soon as we do not observe this clustering any more, but a large eigenvalue ratio of the perturbed eigenvalues, we are in the systematic spectrum.

Theorem 9. Estimator for number of factors
Assume Assumption 1 holds and $O\left(\frac{N}{M}\right) \leq O(1)$. Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. Choose a slowly increasing sequence $g(N,M)$ such that $\frac{g(N,M)}{N} \to 0$ and $g(N,M) \to \infty$. Define perturbed eigenvalues

$$\hat{\lambda}_k = \lambda_k + g(N,M)$$

and the perturbed eigenvalue ratio statistics:

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \textit{for } k = 1, ..., N-1$$

Define

$$\hat{K}(\gamma) = \max\{k \leq N-1 : ER_k > 1+\gamma\}$$

for $\gamma > 0$*. If* $ER_k < 1+\gamma$ *for all* k*, then set* $\hat{K}(\gamma) = 0$*. Then for any* $\gamma > 0$

$$\hat{K}(\gamma) \xrightarrow{p} K.$$

Assume in addition that Assumption 3 holds. Denote the ordered eigenvalues of $\hat{X}^{C\top}\hat{X}^C$ *by* $\lambda_1^C \geq ... \geq \lambda_N^C$ *and analogously for* $\hat{X}^{D\top}\hat{X}^D$ *by* $\lambda_1^D \geq ...\lambda_N^D$*. Define* $\hat{K}^C(\gamma)$ *and* $\hat{K}^D(\gamma)$ *as above but using* λ_i^C *respectively* λ_i^D*. Then for any* $\gamma > 0$

$$\hat{K}^C(\gamma) \xrightarrow{p} K^C \qquad \hat{K}^D(\gamma) \xrightarrow{p} K^D$$

where K^C *is the number of continuous factors and* K^D *is the number of jump factors.*

Some of the most relevant estimators for the number of factors in large-dimensional factor models based on long-horizons are the Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) estimators. The Bai and Ng (2002) paper uses an information criterion, while Onatski applies an eigenvalue difference estimator and Ahn and Horenstein an eigenvalue ratio approach. In simulations the last two estimators seem to perform well.[13] My estimator combines elements of the Ahn and Horenstein estimator as I analyze eigenvalue ratios and elements of the Onatski estimator as I use a clustering argument. In contrast to these two approaches my results are not based on random matrix theory. Under the strong assumptions of random matrix theory a certain fraction of the small eigenvalues will be bounded from below and above and the largest residual eigenvalues will cluster. Onatksi analyses the difference in eigenvalues. As long as the eigenvalue difference is small, it is likely to be part of the residual spectrum because of the clustering effect. The first time the eigenvalue difference is above a threshold, it indicates the beginning of the systematic spectrum. The Ahn and Horenstein method looks for the maximum in the eigenvalue ratios. As the smallest systematic eigenvalue is unbounded, while up to a certain index the nonsystematic eigenvalues are bounded from above and below, consistency follows. However, if the first systematic factor is stronger than the other weak systematic factors the Ahn and Horenstein method can fail in simulations with realistic values.[14]

[13] See for example the numerical simulations in Onatski (2010) and Ahn and Horenstein (2013).

[14] Their proposal to demean the data which is essentially the same as projecting out an equally weighted market

In this sense the clustering argument of Onatksi is more appealing as it focusses on the residual spectrum and tries to identify when the spectrum is unlikely to be due to residual terms. For the same reason my perturbed eigenvalue ratio estimator performs well in simulations with strong and weak factors.

My estimator depends on two choice variables: the perturbation g and the cutoff γ. In contrast to Bai and Ng, Onatski or Ahn and Horenstein we do not need to choose some upper bound on the number of factors. Although consistency follows for any g or γ satisfying the necessary conditions, the finite sample properties will obviously depend on them. As a first step for understanding the factor structure I recommend plotting the perturbed eigenvalue ratio statistic. In all my simulations the transition from the idiosyncratic spectrum to the systematic spectrum is very apparent. Based on simulations a good choice for the perturbation is $g = \sqrt{N} \cdot median(\{\lambda_1, ..., \lambda_N\})$. Obviously this choice assumes that the median eigenvalue is bounded from below, which is not guaranteed by our assumptions but almost always satisfied in practice. In the simulations I also test different specifications for g, e.g. $\log(N) \cdot median(\{\lambda_1, ..., \lambda_N\})$. My estimator is very robust to the choice of the perturbation value. A more delicate issue is the cutoff γ. Simulations suggest that γ between 0.05 and 0.2 performs very well. As we are actually only interested in detecting a deviation from clustering around 1, we can also define $1 + \gamma$ to be proportional to a moving average of perturbed eigenvalue ratios.

What happens if we employ my eigenvalue ratio estimator with a constant perturbation or no perturbation at all? Under stronger assumptions on the idiosyncratic processes, the eigenvalue ratio estimator is still consistent as Proposition 1 shows:

Proposition 1. Onatski-type estimator for number of factors
Assume Assumptions 1 and 3 hold and $\frac{N}{M} \to c > 0$. *In addition assume that*

1. *The idiosyncratic terms follow correlated Brownian motions:*

$$e(t) = A\epsilon(t)$$

   where $\epsilon(t)$ *is a vector of* N *independent Brownian motions.*
2. *The correlation matrix* A *satisfies:*
   (a) *The eigenvalue distribution function* $\mathcal{F}^{AA^\top}$ *converges to a probability distribution function* $\mathcal{F}_A$.
   (b) *The distribution* $\mathcal{F}_A$ *has bounded support,* $u(\mathcal{F}) = min(z : \mathcal{F}(z) = 1)$ *and* $u(\mathcal{F}^{AA^\top}) \to u(\mathcal{F}_A) > 0$.
   (c) $\liminf_{z \to 0} z^{-1} \int_{u(\mathcal{F}_A)-z}^{u(\mathcal{F}_A)} d\mathcal{F}_A(\lambda) = k_A > 0$.

portfolio does not perform well in simulations with a strong factor. The obvious extension to project out the strong factors does also not really solve the problem as it is unclear how many factors we have to project out.

Denote the ordered eigenvalues of $X^\top X$ *by* $\lambda_1 \geq ... \geq \lambda_N$*. Define*

$$\hat{K}^{ON}(\gamma) = \max\left\{k \leq K^{ON}_{\max} : \frac{\lambda_k}{\lambda_{k+1}} \geq \gamma\right\}$$

for any $\gamma > 0$ *and slowly increasing sequence* $K^{ON}_{\max}$ *s.t.* $\frac{K^{ON}_{\max}}{N} \to 0$*. Then*

$$\hat{K}^{ON}(\gamma) \xrightarrow{p} K.$$

Under the Onatski assumptions in Proposition 1, we could also set $g = C$ to some constant, which is independent of N and M. We would get

$$\begin{aligned} ER_K &= O_p(N) \\ ER_k &= \frac{\lambda_k + C}{\lambda_{k+1} + C} \xrightarrow{p} 1 \qquad k \in [K+1, K^{ON}_{\max}]. \end{aligned}$$

However, the Onatski-type estimator in Proposition 1 fails if we use the truncated data $\hat{X}^C$ or $\hat{X}^D$. Proposition 1 shows that Theorem 9 is in some sense robust to the perturbation if we are willing to make stronger assumptions. The stronger assumptions are needed to use results from random matrix theory to obtain a clustering in the residual spectrum.

7 Microstructure noise

While my estimation theory is derived under the assumption of synchronous data with negligible microstructure noise, I extend the model to estimate the effect of microstructure noise on the spectrum of the factor estimator. Inference on the volatility of a continuous semimartingale under noise contamination can be pursued using smoothing techniques. Several approaches have been developed, prominent ones by Aït-Sahalia and Zhang (2005b), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009) in the one-dimensional setting and generalizations for a noisy non-synchronous multi-dimensional setting by Aït-Sahalia et al. (2010), Podolskij and Vetter (2009), Barndorff-Nielsen et al. (2011), Zhang (2011) and Bibinger and Winkelmann (2014) among others. However, neither the microstructure robust estimators nor the non-synchronicity robust estimators can be easily extended to our large dimensional problem. It is beyond the scope of this paper to develop the asymptotic theory for these more general estimators in the context of a large dimensional factor model and I leave this to future research.

The main results of my paper assume synchronous data with negligible microstructure noise. Using for example 5-minute sampling frequency as commonly advocated in the literature on realized volatility estimation, e.g. Andersen et al. (2001) and the survey by Hansen and Lunde (2006), seems to justify this assumption and still provides enough high-frequency observations to apply my estimator to a monthly horizon.

Here I extend the model and show how the microstructure noise affects the largest eigenvalue of the residual matrix. The estimation of the number of factors crucially depends on the size of this largest idiosyncratic eigenvalue. This theorem can be used to show that the estimator for the number of factors does not change in the presence of microstructure noise. It can also be used to derive an estimator for the variance of the microstructure noise. This is the first estimator for the variance of microstructure noise that uses the information in a large cross-section. If we do not use microstructure noise robust estimators for the quadratic covariation matrix, the usual strategy is to use a lower sampling frequency that trades off the noise bias with the estimation variance. This theorem can provide some guidance if the frequency is sufficiently low to neglect the noise.

Theorem 10. Upper bound on impact of noise
Assume we observe the true asset price with noise:

$$Y_i(t_j) = X_i(t_j) + \tilde{\epsilon}_{j,i}$$

where the noise $\tilde{\epsilon}_{j,i}$ *is i.i.d.* $(0, \sigma_\epsilon^2)$ *and independent of* X *and has finite fourth moments. Furthermore assume that Assumption 1 holds and that* $\frac{N}{M} \to c < 1$. *Denote increments of the noise by* $\epsilon_{j,i} = \tilde{\epsilon}_{j+1,i} - \tilde{\epsilon}_{j,i}$. *Then we can bound the impact of noise on the largest eigenvalue of the idiosyncratic spectrum:*

$$\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \min_{s\in[K+1,N-K]}\left(\lambda_s\left(\frac{Y^\top Y}{N}\right)\frac{1}{1+cos\left(\frac{s+r+1}{N}\pi\right)}\right) \cdot 2\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2 + o_p(1).$$

The variance of the microstructure noise is bounded by

$$\sigma_\epsilon^2 \leq \frac{c}{2(1-\sqrt{c})^2}\min_{s\in[K+1,N-K]}\left(\lambda_s\left(\frac{Y^\top Y}{N}\right)\frac{1}{1+cos\left(\frac{s+r+1}{N}\pi\right)}\right) + o_p(1)$$

where $\lambda_s\left(\frac{Y^\top Y}{N}\right)$ *denotes the sth largest eigenvalue of a symmetric matrix* $\frac{Y^\top Y}{N}$.

Remark 1. *For* $s = \frac{1}{2}N - K - 1$ *the inequality simplifies to*

$$\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \lambda_{1/2N-K-1}\left(\frac{Y^\top Y}{N}\right) \cdot 2\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2 + o_p(1)$$

respectively

$$\sigma_\epsilon^2 \leq \frac{c}{2(1-\sqrt{c})^2} \cdot \lambda_{1/2N-K-1}\left(\frac{Y^\top Y}{N}\right) + o_p(1).$$

Hence, the contribution of the noise on the largest eigenvalue of the idiosyncratic part and the microstructure noise variance can be bounded by approximately the median eigenvalue of the observed quadratic covariation matrix multiplied by a constant that depends only on the ratio of M and N.

8 Identifying the Factors

This section develops a new estimator for testing if a set of estimated statistical factors is the same as a set of observable economic variables. As I have already noted before, factor models are only identified up to invertible transformations. Two sets of factors represent the same factor model if the factors span the same vector space. When trying to interpret estimated factors by comparing them with economic factors, we need a measure to describe how close two vector spaces are to each other. As proposed by Bai and Ng (2006) the generalized correlation is a natural candidate measure. Let F be our K-dimensional set of factor processes and G be a K_G-dimensional set of economic candidate factor processes. We want to test if a linear combination of the candidate factors G can replicate some or all of the true factors F. The first generalized correlation is the highest correlation that can be achieved through a linear combination of the factors F and the candidate factors G. For the second generalized correlation we first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest possible correlation that can be achieved through linear combinations of the remaining $K-1$ respectively K_G-1 dimensional subspaces. This procedure continues until we have calculated the $min(K, K_G)$ generalized correlation. Mathematically the generalized correlations are the square root of the $min(K, K_G)$[15] largest eigenvalues of the matrix $[F,G]^{-1}[F,F][G,G]^{-1}[G,F]$. If $K = K_G = 1$ it is simply the correlation as measured by the quadratic covariation. If for example for $K = K_G = 3$ the generalized correlations are $\{1,1,0\}$ it implies that there exists a linear combination of the three factors in G that can replicate two of the three factors in F.[16] I show that under general conditions the estimated factors $\hat{F}$, $\hat{F}^C$ and $\hat{F}^D$ can be used instead of the true unobserved factors.

[15] Using $min(K, K_G)$ instead of $max(K, K_G)$ is just a labeling convention. All the generalized correlations after $min(K, K_G)$ are zero and hence usually neglected.

[16] Although labeling the measure as a correlation, we do not demean the data. This is because the drift term essentially describes the mean of a semimartingale and when calculating or estimating the quadratic covariation it is asymptotically negligible. Hence, the generalized correlation measure is based only on inner products and the generalized correlations correspond to the singular values of the matrix $[F,G]$ if F and G are orthonormalized with respect to the inner product $[.,.]$.

Unfortunately, in this high-frequency setting there does not seem to exist a theory for confidence intervals for the individual generalized correlations.[17] It is well-known that if F and G are observed and *i.i.d.* normally distributed then $\frac{\sqrt{M}(\hat{\rho}_k^2-\rho_k^2)}{2\rho_k(1-\rho_k^2)} \xrightarrow{D} N(0,1)$ for $k = 1, ..., \min(K_F, K_G)$ where ρ_k is the kth generalized correlation.[18]. The result can also be extended to elliptical distributions. However, the normalized increments of stochastic processes that can realistically model financial time series are neither normally nor elliptically distributed. Hence, we cannot directly make use of these results as for example in Bai and Ng (2006). However, I have developed an asymptotic distribution theory for the sum of squared generalized correlations, which I label as total generalized correlation. With the total generalized correlation we can test if a set of economic factors represents the same factor model as the statistical factors.

The total generalized correlation denoted by $\bar{\rho}$ is defined as the sum of the squared generalized correlations $\bar{\rho} = \sum_{k=1}^{\min(K_F,K_G)} \rho_k^2$. It is equal to

$$\bar{\rho} = trace\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F]\right).$$

The estimator for the total generalized correlation is defined as

$$\hat{\bar{\rho}} = trace\left((\hat{F}^\top \hat{F})^{-1}(\hat{F}^\top G)(G^\top G)^{-1}(G^\top \hat{F})\right).$$

As the trace operator is a differentiable function and the quadratic covariation estimator is asymptotically mixed-normally distributed we can apply a delta method argument to show that $\sqrt{M}(\hat{\bar{\rho}} - \bar{\rho})$ is asymptotically mixed-normally distributed as well.

A test for equality of two sets tests if $\bar{\rho} = \min(K_F, K_G)$. As an example consider $K_F = K_G = 3$ and the total generalized correlation is equal to 3. In this case $F(t)$ is a linear transformation of $G(t)$ and both describe the same factor model. Based on the asymptotic normal distribution of $\hat{\bar{\rho}}$ we can construct a test statistic and confidence intervals. The null hypothesis is $\bar{\rho} = \min(K_F, K_G)$.

In the simple case of $K_F = K_G = 1$ the squared generalized correlation and hence also the total generalized correlation correspond to a measure of R^2, i.e. it measures the amount of variation that is explained by G_1 in a regression of F_1 on G_1. My measure of total generalized correlations can be interpreted as a generalization of R^2 for a regression of a vector space on another vector space.

Theorem 11. Asymptotic distribution for total generalized correlation
Assume $F(t)$ is a factor process as in Assumption 1. Denote by $G(t)$ a K_G-dimensional process satisfying Definition 1. The process G is either (i) a well-diversified portfolio of X, i.e. it can

[17] Aït-Sahalia and Xiu's (2015a) distribution results on the eigenvalues of estimated quadratic covariation matrixes can potentially be extended to close this gap.

[18] See for example Anderson (1984)

be written as $G(t) = \frac{1}{N}\sum_{i=1}^N w_i X_i(t)$ *with* $\|w_i\|$ *bounded for all* i *or (ii)* G *is independent of the residuals* $e(t)$*. Furthermore assume that* $\frac{\sqrt{M}}{N} \to 0$*. The* $M \times K_G$ *matrix of increments is denoted by* G*. Assume that*[19]

$$\sqrt{M}\left(\begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix} - \begin{pmatrix} [F,F] & [F,G] \\ [G,F] & [G,G] \end{pmatrix}\right) \stackrel{L-s}{\to} N(0,\Pi).$$

Then

$$\sqrt{M}\left(\hat{\bar{\rho}} - \bar{\rho}\right) \stackrel{L-s}{\to} N(0,\Xi) \qquad \text{and} \qquad \frac{\sqrt{M}}{\sqrt{\Xi}}\left(\hat{\bar{\rho}} - \bar{\rho}\right) \stackrel{D}{\to} N(0,1)$$

with $\Xi = \xi^\top \Pi \xi$ *and*

$$\xi = vec\left(\begin{pmatrix} -\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F][F,F]^{-1}\right)^\top & [F,F]^{-1}[F,G][G,G]^{-1} \\ [G,G]^{-1}[G,F][F,F]^{-1} & -\left([G,G]^{-1}[G,F][F,F]^{-1}[F,G][G,G]^{-1}\right)^\top \end{pmatrix}\right).$$

Here a I present a feasible test statistic for the estimated continuous factors. A feasible test for the jump factors can also be derived.

Theorem 12. A feasible central limit theorem for the generalized continuous correlation

Assume Assumptions 1 to 3 hold. The process G *is either (i) a well-diversified portfolio of* X*, i.e. it can be written as* $G(t) = \frac{1}{N}\sum_{i=1}^N w_i X_i(t)$ *with* $\|w_i\|$ *bounded for all* i *or (ii)* G *is independent of the residuals* $e(t)$*. Furthermore assume that* $\frac{\sqrt{M}}{N} \to 0$*. Denote the threshold estimators for the continuous factors as* $\hat{F}^C$ *and for the continuous component of* G *as* $\hat{G}^C$*. The total generalized continuous correlation is*

$$\bar{\rho}^C = trace\left([F^C,F^C]^{-1}[F^C,G^C][G^C,G^C]^{-1}[G^C,F^C]\right)$$

and its estimator is

$$\hat{\bar{\rho}}^C = trace\left((\hat{F^C}^\top \hat{F}^C)^{-1}(\hat{F^C}^\top \hat{G}^C)(\hat{G^C}^\top \hat{G}^C)^{-1}(\hat{G^C}^\top \hat{F}^C)\right).$$

[19] As explained in for example Barndorff-Nielsen and Shephard (2004a) the statement should be read as $\sqrt{M}\left(vec\left(\begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix}\right) - vec\left(\begin{pmatrix} [F,F] & [F,G] \\ [G,F] & [G,G] \end{pmatrix}\right)\right) \stackrel{L-s}{\to} N(0,\Pi)$, where *vec* is the vectorization operator. Inevitably the matrix Π is singular due to the symmetric nature of the quadratic covariation. A proper formulation avoiding the singularity uses *vech* operators and elimination matrices (See Magnus (1988)).

Then

$$\frac{\sqrt{M}}{\sqrt{\hat{\Xi}^C}}\left(\hat{\bar{\rho}}^C - \bar{\rho}^C\right) \xrightarrow{D} N(0,1)$$

Define the $M \times (K_F + K_G)$ *matrix* $Y = \begin{pmatrix} \hat{F}^C & \hat{G}^C \end{pmatrix}$. *Choose a sequence satisfying* $k \to \infty$ *and* $\frac{k}{M} \to 0$ *and estimate spot volatilities as*

$$\hat{v}_j^{i,r} = \frac{M}{k}\sum_{l=1}^{k-1} Y_{j+l,i}Y_{j+l,r}.$$

The estimator of the $(K_F + K_G) \times (K_F + K_G)$ *quarticity matrix* $\hat{\Pi}^C$ *has the elements*

$$\hat{\Pi}^C_{r+(i-1)(K_F+K_G),n+(m-1)(K_F+K_G)} = \frac{1}{M}\left(1 - \frac{2}{k}\right)\sum_{j=1}^{M-k+1}\left(v_j^{i,r}v_j^{m,n} + v_j^{i,n}v_j^{r,m}\right)$$

for $i, r, m, n = 1, ..., K_F + K_G$. *Estimate* $\hat{\xi}^C = vec(S)$ *for the matrix* S *with block elements*

$$\begin{aligned}
S_{1,1} &= -\left(\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\right)^\top\\
S_{1,2} &= \left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\\
S_{2,1} &= \left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\\
S_{2,2} &= -\left(\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\right)^\top.
\end{aligned}$$

The estimator for the covariance of the total generalized correlation estimator is $\hat{\Xi}^C = \hat{\xi}^{C\top}\hat{\Pi}^C\hat{\xi}^C$.

The assumption that G has to be a well-diversified portfolio of the underlying asset space is satisfied by essentially all economic factors considered in practice, e.g. the market factor or the value, size and momentum factors. Hence, practically it does not impose a restriction on the testing procedure. This assumption is only needed to obtain the same distribution theory for the quadratic covariation of G with the estimated factors as with the true factors.

9 Differences to Long-Horizon Factor Models

The estimation approach of my high-frequency factor model can in general not be mapped into Bai's (2003) general long-horizon factor model. After rescaling the increments, we can interpret the quadratic covariation estimator as a sample covariance estimator. However, in contrast to the covariance estimator, the limiting object will be a random variable and the asymptotic distribution results have to be formulated in terms of stable convergence in law, which is stronger than convergence in distribution. Models with jumps have "heavy-tailed rescaled increments" which cannot be accommodated in Bai's (2003) model. In stochastic volatility or stochastic intensity jump models the data is non-stationary. Some of the results in large dimensional factor analysis do not apply to non-stationary data. In contrast to long-horizon factor analysis the asymptotic distribution of my estimators have a mixed Gaussian limit and so will generally have heavier tails than a normal distribution.

I start with a simple case where the high-frequency problem is nested in the long-horizon model. First, I assume that all stochastic processes are Brownian motions:

$$X_T = \begin{pmatrix} \Lambda_{11} & \cdots & \Lambda_{1K} \\ \vdots & \ddots & \vdots \\ \Lambda_{1K} & \cdots & \Lambda_{NK} \end{pmatrix} \begin{pmatrix} W_{F_1}(t) \\ \vdots \\ W_{F_K}(t) \end{pmatrix} + \begin{pmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{NN} \end{pmatrix} \begin{pmatrix} W_{e_1}(t) \\ \vdots \\ W_{e_N}(t) \end{pmatrix}$$

where all Brownian motions W_{F_k} and W_{e_i} are independent of each other. In this case the quadratic covariation equals

$$[X,X] = \Lambda[F,F]\Lambda^\top + [e,e] = \Lambda\Lambda^\top T + \begin{pmatrix} \sigma_{11}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{NN}^2 \end{pmatrix} T$$

Under standard assumptions $\Lambda\Lambda^\top$ is a $N \times N$ matrix of rank K and its eigenvalues will go to infinity for $N \to \infty$. On the other hand $[e,e]$ has bounded eigenvalues. The problem is the estimation of the unobserved quadratic covariation matrix $[X,X]$ for large N. Although, we can estimate each entry of the matrix with a high precision, the estimation errors will sum up to a non negligible quantity if N is large. In the case of a large-dimensional sample covariance matrix Bai (2003) has solved the problem. If we divide the increments by the square root of the length of the time increments $\Delta_M = T/M$, we end up with a conventional covariance estimator:

$$\sum_{j=1}^{M} (\Delta_j X_i)^2 = \frac{T}{M} \sum_{j=1}^{M} \left(\frac{\Delta_j X_i}{\sqrt{\Delta_M}} \right)^2 \qquad \text{with } \frac{\Delta_j X_i}{\sqrt{\Delta_M}} \sim i.i.d.\ N(0, \Lambda_i \Lambda_i^\top + \sigma_{ii}^2).$$

These rescaled increments satisfy all the assumptions of Bai (2003)'s estimator.

However, for general stochastic process we violate the assumptions in Bai's paper. Assume that the underlying stochastic processes have stochastic volatility and jumps. Both are features that are necessary to model asset prices realistically.

$$F(t) = \int_0^t \sigma_F(s) dW_F(s) + \sum_{s \leq t} \Delta F(s) \qquad e(t) = \int_0^t \sigma_e(s) dW_e(s) + \sum_{s \leq t} \Delta e(s).$$

First, if X_i is allowed to have jumps, then it is easy to show that the rescaled increments $\frac{\Delta_j X_i}{\sqrt{\Delta_M}}$ do not have fourth moments. However, Bai (2003) requires the random variables to have at least 8 moments.[20] Second, the quadratic covariation matrices evaluated at time T will now be random variables given by[21]

$$[F, F] = \int_0^T \sigma_F^2(s) ds + \sum_{s \leq T} \Delta F^2(s) \qquad [e_i, e_k] = \int_0^T \sigma_{e_{i,k}}(s) ds + \sum_{s \leq T} \Delta e_i(s) \Delta e_k(s).$$

and $[X, X] = \Lambda [F, F] \Lambda^\top + [e, e]$. The high-frequency estimator is based on path-wise arguments for the stochastic processes, while Bai's estimator is based on population assumptions. Third, the mode of convergence is now stable convergence in law, which is stronger than simple convergence in distribution.[22] Although the estimator for the quadratic covariation is $\sqrt{M}$ consistent, it has now an asymptotic mixed-Gaussian law:

$$\sqrt{M} \sum_{j=1}^M F_j e_{ji} \stackrel{L-s}{\rightarrow} N\left(0, \int_0^T \sigma_F^2 \sigma_{e_i}^2 ds + \sum_{s \leq T} \Delta F^2(s) \sigma_{e_i}^2(s) + \sum_{s' \leq T} \Delta e_i^2(s') \sigma_F^2(s')\right).$$

This directly affects the distribution of the loadings estimator. Similar arguments apply to the factor estimator.

10 Simulations

This section considers the finite sample properties of my estimators through Monte-Carlo simulations. In the first subsection I use Monte-Carlo simulations to analyze the distribution of my estimators for the loadings, factors and common components. In the second subsection I provide a simulation study of the estimator for the number of factors and compare it to the most popular estimators in the literature.

My benchmark model is a Heston-type stochastic volatility model with jumps. In the general

[20] Assumption C in Bai (2003)

[21] Here I assume that there is only one factor, i.e. $K = 1$.

[22] Assumption F in Bai (2003).

case I assume that the K factors are modeled as

$$dF_k(t) = (\mu - \sigma_{F_k}^2(t))dt + \rho_F \sigma_{F_k}(t) dW_{F_k}(t) + \sqrt{1-\rho_F^2}\sigma_{F_k}(t) d\tilde{W}_{F_k}(t) + J_{F_k} dN_{F_k}(t)$$
$$d\sigma_{F_k}^2(t) = \kappa_F \left(\alpha_F - \sigma_{F_k}^2(t)\right) dt + \gamma_F \sigma_{F_k}(t) d\tilde{W}_{F_k}(t)$$

and the N residual processes as

$$de_i(t) = \rho_e \sigma_{e_i}(t) dW_{e_i}(t) + \sqrt{1-\rho_e^2}\sigma_{e_i}(t) d\tilde{W}_{e_i}(t) + J_{e_i} dN_{e_i}(t) - \mathbb{E}[J_{e_i}]\nu_e dt$$
$$d\sigma_{e_i}^2(t) = \kappa_e \left(\alpha_e - \sigma_{e_i}^2(t)\right) dt + \gamma_e \sigma_{e_i}(t) d\tilde{W}_{e_i}(t)$$

The Brownian motions $W_F, \tilde{W}_F, W_e, \tilde{W}_e$ are assumed to be independent. I set the parameters to values typically used in the literature: $\kappa_F = \kappa_e = 5$, $\gamma_F = \gamma_e = 0.5$, $\rho_F = -0.8$, $\rho_e = -0.3$, $\mu = 0.05$, $\alpha_F = \alpha_e = 0.1$. The jumps are modeled as a compound Poisson process with intensity $\nu_F = \nu_e = 6$ and normally distributed jumps with $J_{F_k} \sim N(-0.1, 0.5)$ and $J_{e_i} \sim N(0, 0.5)$. The time horizon is normalized to $T = 1$.

In order to separate continuous from discontinuous movements I use the threshold $3\hat{\sigma}_X(j)\Delta_M^{0.48}$. The spot volatility is estimated using Barndorff-Nielsen and Shephard's (2006) bi-power volatility estimator on a window of $\sqrt{M}$ observations. Under certain assumptions the bi-power estimator is robust to jumps and estimates the volatility consistently.

In order to capture cross-sectional correlations I formulate the dynamics of X as

$$X(t) = \Lambda F(t) + Ae(t)$$

where the matrix A models the cross-sectional correlation. If A is an identity matrix, then the residuals are cross-sectionally independent. The empirical results suggest that it is very important to distinguish between strong and weak factors. Hence the first factor is multiplied by the scaling parameter $\sigma_{dominant}$. If $\sigma_{dominant} = 1$ then all factors are equally strong. In practice, the first factor has the interpretation of a market factor and has a significantly larger variance than the other weaker factors. Hence, a realistic model with several factors should set $\sigma_{dominant} > 1$.

The loadings Λ are drawn from independent standard normal distributions. All Monte-Carlo simulations have 1000 repetitions. I first simulate a discretized model of the continuous time processes with 2000 time steps representing the true model and then use the data which is observed on a coarser grid with $M = 50, 100$, 250 or 500 observations. My results are robust to changing the number of Monte-Carlo simulations or using a finer time grid for the "true" process.

10.1 Asymptotic Distribution Theory

In this subsection I consider only one factor in order to assess the properties of the limiting distribution, i.e. $K = 1$ and $\sigma_{dominant} = 1$. I consider three different cases:

1. **Case 1: Benchmark model with jumps.** The correlation matrix A is a Toplitz matrix with parameters $(1, 0.2, 0.1)$, i.e. it is a symmetric matrix with diagonal elements 1 and the first two off-diagonals have elements 0.2 respectively 0.1.
2. **Case 2: Benchmark model without jumps.** This model is identical to case 1 but without the jump component in the factors and residuals.
3. **Case 3: Toy model.** Here all the stochastic processes are standard Brownian motions

$$X(t) = \Lambda W_F(t) + W_e(t)$$

After rescaling case 3 is identical to the simulation study considered in Bai (2003).

Obviously, we can only estimate the continuous and jump factors in case 1.

In order to assess the accuracy of the estimators I calculate the correlations of the estimator for the loadings and factors with the true values. If jumps are included, we have additionally correlations for the continuous and jump estimators. In addition for $t = T$ and $i = N/2$ I calculate the asymptotic distribution of the rescaled and normalized estimators:

$$\begin{aligned}
CLT_C &= \left(\frac{1}{N}\hat{V}_{T,i} + \frac{1}{M}\hat{W}_{T,i}\right)^{-1/2} \left(\hat{C}_{T,i} - C_{T,i}\right) \\
CLT_F &= \sqrt{N}\hat{\Theta}_F^{-1/2}(\hat{F}_T - H^{-1}F_T) \\
CLT_\Lambda &= \sqrt{M}\hat{\Theta}_{\Lambda,i}^{-1/2}(\hat{\Lambda}_i - H^\top \Lambda_i)
\end{aligned}$$

Table 1 reports the mean and standard deviation of the correlation coefficients between $\hat{F}_T$ and F_T and $\hat{\Lambda}_i$ and Λ_i based on 1000 simulations. In case 1 I also estimate the continuous and jump part. The correlation coefficient can be considered as a measure of consistency. For the factor processes the correlation is based on the quadratic covariation between the true and the estimated processes. I run the simulations for four combinations of N and M: $N = 200, M = 250$, $N = 100, M = 100$, $N = 500, M = 50$ and $N = 50, M = 500$. The correlation coefficients in all cases are very close to one, indicating that my estimators are very precise. Note, that we can only estimate the continuous and jump factor up to a finite variation part. However, when calculating the correlations, the drift term is negligible. For a small number of high-frequency observations M the continuous and the jump factors are estimated with a lower precision as the total factor. This is mainly due to an imprecision in the estimation of the jumps. In all cases the loadings can be estimated very precisely. The simpler the processes, the better the estimators work. For sufficiently large N and M, increasing M

	N=200, M=250					N=100, M=100				
	Case 1			Case 2	Case 3	Case 1			Case 2	Case 3
	Total	Cont.	Jump			Total	Cont.	Jump		
Corr. F_T	0.994	0.944	0.972	0.997	0.997	0.986	0.789	0.943	0.994	0.997
SD F_T	0.012	0.065	0.130	0.001	0.000	0.037	0.144	0.165	0.002	0.000
Corr. Λ	0.995	0.994	0.975	0.998	0.998	0.986	0.966	0.949	0.994	0.998
SD Λ	0.010	0.008	0.127	0.001	0.000	0.038	0.028	0.157	0.002	0.000
	N=500, M=50					N=50, M=500				
	Case 1			Case 2	Case 3	Case 1			Case 2	Case 3
	Total	Cont.	Jump			Total	Cont.	Jump		
Corr. F_T	0.997	0.597	0.926	0.999	0.999	0.973	0.961	0.954	0.988	0.990
SD F_T	0.006	0.196	0.151	0.001	0.000	0.067	0.028	0.141	0.005	0.002
Corr. Λ	0.979	0.921	0.906	0.987	0.990	0.991	0.997	0.974	0.999	0.999
SD Λ	0.027	0.051	0.175	0.005	0.002	0.053	0.002	0.128	0.001	0.000

Table 1: Mean and standard deviations of estimated correlation coefficients between $\hat{F}_T$ and F_T and $\hat{\Lambda}_i$ and Λ_i based on 1000 simulations.

improves the estimator for the loadings, while increasing N leads to a better estimation of the factors. Overall, the finite sample properties for consistency are excellent.

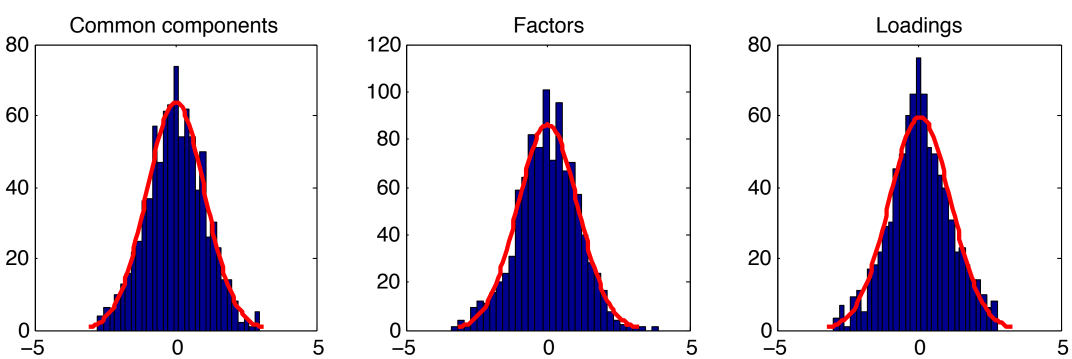

Figure 1: Case 1 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

Table 2 and Figures 1 to 3 summarize the simulation results for the normalized estimators CLT_C, CLT_F and CLT_Λ. The asymptotic distribution theory suggests that they should be $N(0,1)$ distributed. The tables list the means and standard deviations based on 1000 simulations. For the toy model in case 3 the mean is close to 0 and the standard deviation almost 1, indicating that the distribution theory works. Figure 3 depicts the histograms overlaid with a normal distribution. The asymptotic theory provides a very good approximation to the fi-

N=200,	M=250	CLT_C	CLT_F	CLT_Λ	N=100,	M=100	CLT_C	CLT_F	CLT_Λ
Case 1	Mean	0.023	0.015	0.051	Case 1	Mean	-0.047	0.025	-0.006
	SD	1.029	1.060	1.084		SD	0.992	1.139	1.045
Case 2	Mean	0.004	-0.007	-0.068	Case 2	Mean	-0.005	0.030	0.041
	SD	1.040	1.006	1.082		SD	1.099	1.046	1.171
Case 3	Mean	0.000	0.002	0.003	Case 3	Mean	0.024	-0.016	-0.068
	SD	1.053	1.012	1.049		SD	1.039	1.060	1.091
N=500,	M=50	CLT_C	CLT_F	CLT_Λ	N=50,	M=500	CLT_C	CLT_F	CLT_Λ
Case 1	Mean	-0.026	-0.012	-0.029	Case 1	Mean	-0.005	-0.044	0.125
	SD	0.964	1.308	1.002		SD	1.055	4.400	1.434
Case 2	Mean	-0.028	-0.009	0.043	Case 2	Mean	0.012	-0.018	-0.020
	SD	1.120	1.172	1.178		SD	0.989	1.038	1.178
Case 3	Mean	-0.064	0.003	0.018	Case 3	Mean	0.053	0.030	-0.013
	SD	1.079	1.159	1.085		SD	1.015	1.042	1.141

Table 2: Mean and standard deviation of normalized estimators for the common component, factors and loadings based on 1000 simulations

nite sample distributions. Adding stochastic volatility and weak cross-sectional correlation still provides a good approximation to a normal distribution. The common component estimator is closer to the asymptotic distribution than the factor or loading estimator. Even in case 1 with the additional jumps the approximation works well. The common component estimator still performs the best. Without an additional finite sample correction the loading estimator in case 1 would have some large outliers. In more detail, the derivations for case 1 assume that the time increments are sufficiently small such that the two independent processes $F(t)$ and $e_i(t)$ do not jump during the same time increment. Whenever this happens the rescaled loadings statistic explodes. For very few of the 1000 simulations in case 1 we observe this problem and exclude these simulations. I have set the length of the local window in the covariance estimation of the loadings estimator to $k = \sqrt{M}$. The estimator for the covariance of the factors assumes cross-sectional independence, which is violated in the simulation example as well as Assumption 5. Nevertheless in the simulations the normalized statistics approximate a normal distribution very well. Overall, the finite sample properties for the asymptotic distribution work well.

10.2 Number of Factors

In this subsection I analyze the finite sample performance of my estimator for the number of factors and show that it outperforms or is at least as good as the most popular estimators in the literature. One of the main motivations for developing my estimator is that the assumptions needed for the Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) estimator

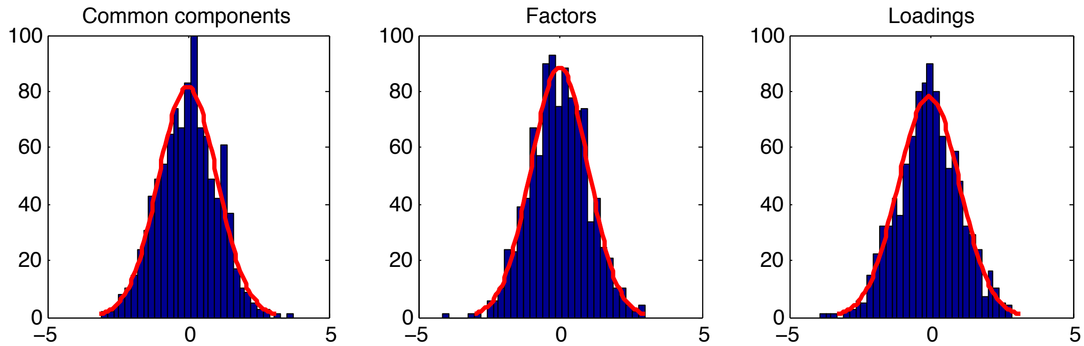

Figure 2: Case 2 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

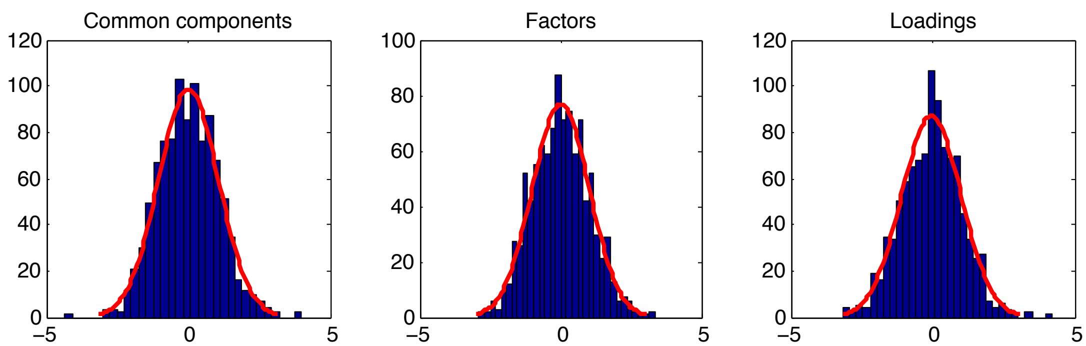

Figure 3: Case 3 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

cannot be extended to the general processes that we need to consider. In particular all three estimators assume essentially that the residuals can be written in the form BEA, where B is a $T \times T$ matrix capturing serial correlation, A is a $N \times N$ matrix modeling the cross-sectional correlation and E is a $T \times N$ matrix of i.i.d. random variables with finite fourth moments. Such a formulation rules out jumps and a complex stochastic volatility structure.

In the first part of this section we work with a variation of the toy model such that we can apply all four estimators and compare them:

$$X(t) = \Lambda W_F(t) + \theta A W_e(t)$$

where all the Brownian motions are independent and the $N \times N$ matrix A models the cross-sectional dependence, while θ captures the signal-to-noise ratio. The matrix A is a Toplitz matrix with parameters $(1, a, a, a, a^2)$, i.e. it is a symmetric matrix with diagonal element 1 and the first four off-diagonals having the elements a, a, a and a^2. A dominant factor is modeled with $\sigma_{dominant} > 1$. Note, that after rescaling this is the same model that is also considered in Bai and Ng, Onatski and Ahn and Horenstein. Hence, these results obviously extend to the long horizon framework. In the following simulations we always consider three factors, i.e. $K = 3$.

I simulate four scenarios:

1. Scenario 1: Dominant factor, large noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = \sqrt{10}$, $\theta = 6$ and $a = 0.5$.
2. Scenario 2: No dominant factor, large noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = 1$, $\theta = 6$ and $a = 0.5$.
3. Scenario 3: No dominant factor, small noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = 1$, $\theta = 1$ and $a = 0.5$.
4. Scenario 4: Toy model $\sigma_{dominant} = 1$, $\theta = 1$ and $a = 0$.

My empirical studies in Pelger (2015) suggest that in the data the first systematic factor is very dominant with a variance that is 10 times larger then those of the other weaker factors. Furthermore the idiosyncratic part seems to have a variance that is at least as large as the variance of the common components. Both findings indicate that scenario 1 is the most realistic case and any estimator of practical relevance must also work in this scenario.

My perturbed eigenvalue ratio statistic has two choice parameters: the perturbation $g(N, M)$ and the cutoff γ. In the simulations I set the cutoff equal to $\gamma = 0.2$. For the perturbation I consider the two choices $g(N, M) = \sqrt{N} \cdot median\{\lambda_1, ..., \lambda_N\}$ and $g(N, M) = \log(N) \cdot median\{\lambda_1, ..., \lambda_N\}$. The first estimator is denoted by $ERP1$, while the second is $ERP2$. All my results are robust to these choice variables. The Onatski (2010) estimator is denoted by *Onatski* and I use the same parameters as in his paper. The Ahn and Horenstein (2013) estimator is labeled as *Ahn*. As suggested in their paper, for their estimator I first demean the data in the cross-sectional and time dimension before applying principal component analysis. *Bai* denotes the BIC3 estimator of Bai and Ng (2002). The BIC3 estimator outperforms the other versions of the Bai and Ng estimators in simulations. For the last three estimators, we need to define an upper bound on the number of factors, which I set equal to $k_{max} = 20$. The main results are not affected by changing k_{max}. For $ERP1$ and $ERP2$ we consider the whole spectrum. The figures and plots are based on 1000 simulations.

Obviously there are more estimators in the literature, e.g. Harding (2013), Alessi, Barigozzi and Capasso (2010) and Hallin and Liska (2007). However, the simulation studies in their papers indicate that the Onatski and Ahn and Horenstein estimators dominate most other estimators.

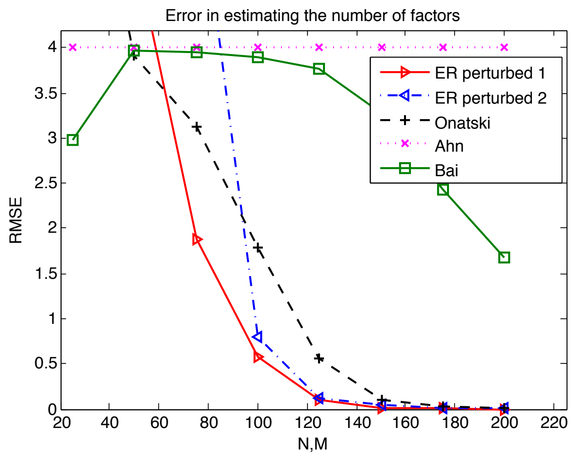

Figure 4: RMSE (root-mean squared error) for the number of factors in scenario 1 for different estimators with $N = M$.

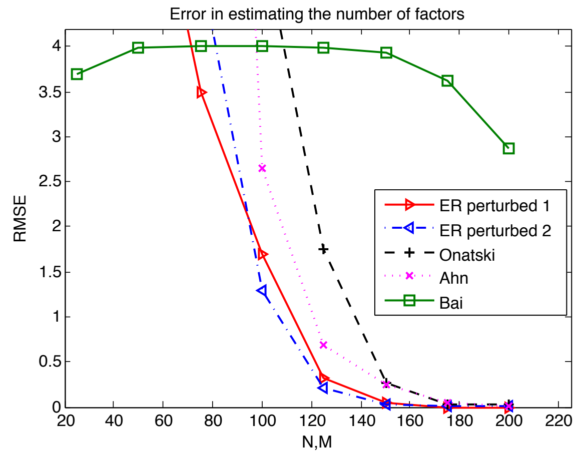

Figure 5: RMSE (root-mean squared error) for the number of factors in scenario 2 for different estimators with $N = M$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.32	0.18	0.49	4.00	3.74
Mean	2.79	2.88	2.76	1.00	1.09
Median	3	3	3	1	1
SD	0.52	0.41	0.66	0.00	0.28
Min	1	1	1	1	1
Max	3	4	5	1	2

Table 3: Scenario 1: $N = M = 125$, $K = 3$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	1.48	0.87	1.99	0.73	3.99
Mean	2.39	2.62	2.31	2.56	1.00
Median	3	3	3	3	1
SD	1.05	0.85	1.23	0.73	0.06
Min	0	0	0	1	1
Max	4	4	6	4	2

Table 4: Scenario 2: $N = M = 125$, $K = 3$.

Figures 4 to 7 plot the root-mean squared error for the different estimators for a growing number $N = M$ and show that my estimators strongly outperform or are at least as good as the other estimators. In the most relevant Scenario 1 depicted in Figure 4 only the $ERP1$, $ERP2$ and *Onatski* estimator are reliable. This is because these three estimators focus on the residual spectrum and are not affected by strong factors. Although we apply the demeaning as proposed in Ahn and Horenstein, their estimator clearly fails. Table 3 shows the summary statistics for this scenario. *Ahn* and *Bai* severely underestimate the number of factors, while the $ERP1$ and $ERP2$ estimators are the best. Note, that the maximal error for both ERP estimators is smaller than for *Onatski*. In Figure 5 we remove the strong factor and the performance of *Ahn* drastically improves. However $ERP1$ and $ERP1$ still show a comparable performance. In the less realistic Scenarios 3 and 4, all estimators are reliable and perform equally well.

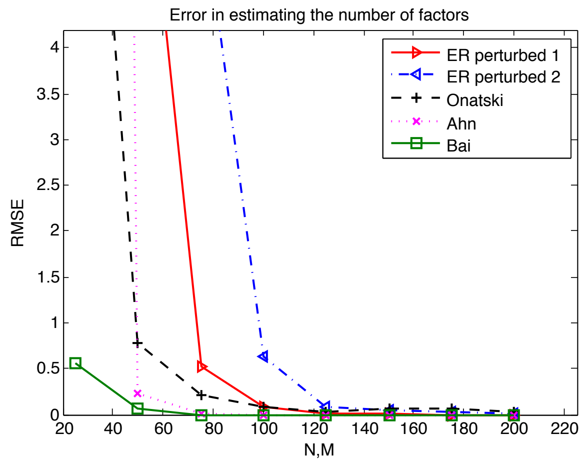

Figure 6: RMSE (root-mean squared error) for the number of factors in scenario 3 for different estimators with $N = M$.

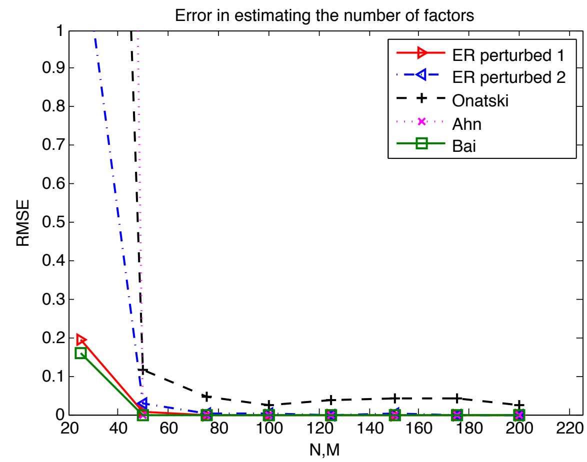

Figure 7: RMSE (root-mean squared error) for the number of factors in scenario 4 for different estimators with $N = M$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.00	0.01	0.06	0.00	0.00
Mean	3.00	3.01	3.03	3.00	3.00
Median	3	3	3	3	3
SD	0.03	0.08	0.24	0.00	0.00
Min	3	3	3	3	3
Max	4	4	7	3	3

Table 5: Scenario 3: $N = M = 125$, $K = 3$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.00	0.00	0.05	0.00	0.00
Mean	3.00	3.00	3.03	3.00	3.00
Median	3	3	3	3	3
SD	0.00	0.03	0.22	0.00	0.00
Min	3	3	3	3	3
Max	3	4	7	3	3

Table 6: Scenario 4: $N = M = 125$, $K = 3$.

Figures 8 and 9 show $ERP1$ applied to the benchmark model Case 1 from the last subsection. The first dominant factor has a continuous and a jump component, while the other two weak factors are purely continuous. Hence, we have $K = 3$, $K^C = 3$, $K^D = 1$ and $\sigma_{dominant} = 3$. I simulate 100 paths for the perturbed eigenvalue ratio and try to estimate K, K^C and K^D. We can clearly see that $ERP1$ clusters for $k > 3$ in the total and continuous case respectively $k > 1$ in the jump case and increases drastically at the true number of factors. How the cutoff threshold γ has to be set, depends very much on the data set. The choice of $\gamma = 0.2$, that worked very well in my previous simulations, would potentially not have been the right choice for Figures 8 and 9. Nevertheless, just by looking at the plots it is very apparent what the right number of factors should be. Therefore, I think plotting the perturbed eigenvalue ratios is a very good first step for understanding the potential factor structure in the data.

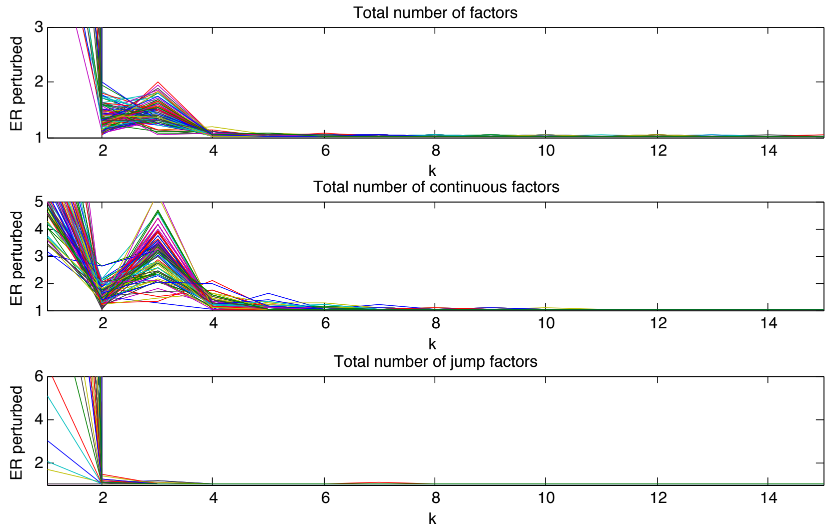

Figure 8: Perturbed eigenvalue ratios (ERP1) in the benchmark case 1 with $K = 3$, $K^C = 3$, $K^D = 1$, $\sigma_{dominant} = 3$, $N = 200$ and $M = 250$ for 100 simulated paths.

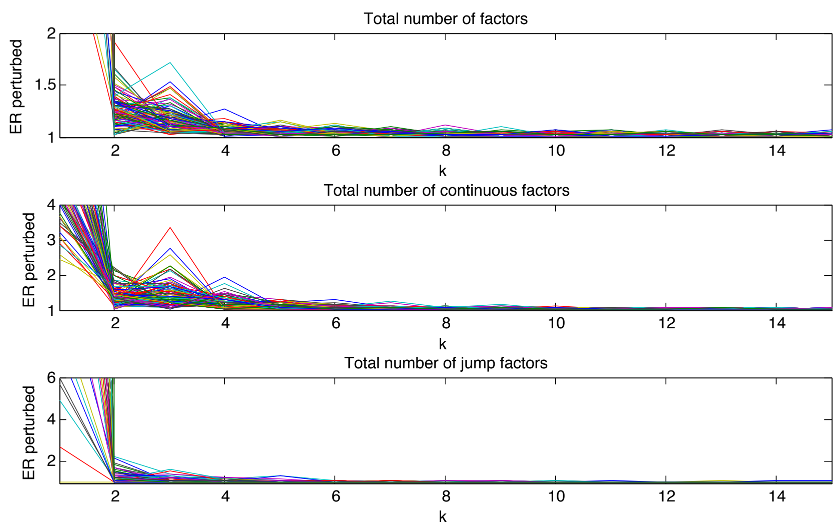

Figure 9: Perturbed eigenvalue ratios (ERP1) in the benchmark case 1 with $K = 3$, $K^C = 3$, $K^D = 1$, $\sigma_{dominant} = 3$, $N = 100$ and $M = 100$ for 100 simulated paths.

11 Conclusion

This paper studies factor models in the new setting of a large cross section and many high-frequency observations under a fixed time horizon. I propose a principal component estimator based on the increments of the observed time series, which is a simple and feasible estimator. For this estimator I develop the asymptotic distribution theory. Using a simple truncation approach the same methodology allows to estimate continuous and jump factors. My results are obtained under very general conditions for the stochastic processes and allow for cross-sectional and serial correlation in the residuals. I also propose a novel estimator for the number of factors, that can also consistently estimate the number of continuous and jump factors under the same general conditions. In two extensions I propose a new test for comparing estimated statistical factors with observed economic factors and a new estimator for the variance of microstructure noise.

In an extensive empirical study in Pelger (2015) I apply the estimation approaches developed in this paper to 5 minutes high-frequency price data of S&P 500 firms from 2003 to 2012. I can show that the continuous factor structure is highly persistent in some years, but there is also time variation in the number and structure of factors over longer horizons. For the time period 2007 to 2012 I estimate four continuous factors which can be approximated very well by a market, oil, finance and electricity factor. The value, size and momentum factors play no significant role in explaining these factors. From 2003 to 2006 one continuous systematic factor disappears. Systematic jump risk also seems to be different from systematic continuous risk. There seems to exist only one persistent jump factor, namely a market jump factor. Using short-maturity, at-the-money implied volatilities from option price data for the same S&P 500 firms from 2003 to 2012 I analyze the systematic factor structure of the volatilities. There there seems to be only one persistent market volatility factor, while during the financial crisis an additional temporary banking volatility factor appears. Based on the estimated factors, I can decompose the leverage effect, i.e. the correlation of the asset return with its volatility, into a systematic and an idiosyncratic component. The negative leverage effect is mainly driven by the systematic component, while the idiosyncratic component can be positively correlated with the volatility. These findings are important as they can rule out popular explanations of the leverage effect, which do not distinguish between systematic and non-systematic risk.

Arbitrage pricing theory links risk premiums to systematic risk. In future projects I want to analyze the ability of the high-frequency factors to price the cross-section of returns. Furthermore I would like to explore the possibility to use even higher sampling frequencies by developing a microstructure noise robust estimation method.

A Structure of Appendix

The appendix is structured as follows. Appendix B specifies the class of stochastic processes used in this paper. In Appendix C I collect some intermediate asymptotic results, which will be used in the subsequent proofs. Appendix D proves the results for the loading estimator. Appendix E treats the estimation of the factors. In Appendix F I show the results for the common components. In Appendix G I derive consistent estimators for the covariance matrices of the estimators. Appendix H deals with separating the continuous and jump factors. The estimation of the number of factors is in Appendix I. Appendix J proves the test for identifying the factors. Last but not least I discuss the proofs for microstructure noise in Appendix K. Finally, for convenience Appendix L contains a collection of limit theorems. In the proofs C is a generic constant that may vary from line to line.

B Assumptions on Stochastic Processes

Definition 1. Locally bounded special Itô semimartingales
The stochastic process Y is a locally bounded special Itô semimartingale if it satisfies the following conditions. Y is a d-dimensional special Itô semimartingale on some filtered space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$, which means it can be written as

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \delta(s,x)(\mu - \nu)(ds, dx)$$

where W is a d-dimensional Brownian motion and μ is a Poisson random measure on $\mathbb{R}_+ \times E$ with $(E, \mathbb{E})$ an auxiliary measurable space on the space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$. The predictable compensator (or intensity measure) of μ is $\nu(ds, dx) = ds \times v(dx)$ for some given finite or sigma-finite measure on $(E, \mathbb{E})$. This definition is the same as for an Itô semimartingale with the additional assumption that $\| \int_0^t \int_E \delta(s,x)\mathbb{1}_{\{\|\delta\|>1\}}\nu(ds,dx)\| < \infty$ for all t. Special semimartingales have a unique decomposition into a predictable finite variation part and a local martingale part.

The coefficients $b_t(\omega), \sigma_t(\omega)$ and $\delta(\omega, t, x)$ are such that the various integrals make sense (see Jacod and Protter (2012) for a precise definition) and in particular b_t and σ_t are optional processes and δ is a predictable function.

The volatility σ_t is also a d-dimensional Itô semimartingale of the form

$$\begin{aligned}\sigma_t =& \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s + \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|\leq 1\}}\tilde{\delta}(s,x)(\mu - \nu)(ds, dx) \\ &+ \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}}\tilde{\delta}(s,x)\mu(ds, dx)\end{aligned}$$

where W' is another Wiener process independent of (W, μ). Denote the predictable quadratic co-variation process of the martingale part by $\int_0^t a_s ds$ and the compensator of $\int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}} \tilde{\delta}(s,x)\mu(ds,dx)$ by $\int_0^t \tilde{a}_s ds$.

1. *I assume a local boundedness condition holds for Y:*
   - *The process b is locally bounded and càdlàg.*
   - *The process σ is càdlàg.*
   - *There is a localizing sequence τ_n of stopping times and, for each n, a deterministic nonnegative function Γ_n on E satisfying $\int \Gamma_n(z)^2 v(dz) < \infty$ and such that $\|\delta(\omega,t,z)\| \wedge 1 \leq \Gamma_n(z)$ for all (ω, t, z) with $t \leq \tau_n(\omega)$.*
2. *The volatility process also satisfy a local boundedness condition:*
   - *The processes $\tilde{b}$ a and $\tilde{a}$ are locally bounded and progressively measurable*
   - *The processes $\tilde{\sigma}$ and $\tilde{b}$ are càdlàg or càglàd*
3. *Furthermore both processes $\sigma\sigma^\top$ and $\sigma_{t-}\sigma_{t-}^\top$ take their values in the set of all symmetric positive definite $d \times d$ matrices.*

More details on high frequency models and asymptotics can be found in the book by Aït-Sahalia and Jacod (2014).

C Some Intermediate Asymptotic Results

C.1 Convergence Rate Results

Proposition C.1. *Assume Y is a d-dimensional Itô-semimartingale satisfying Definition 1:*

$$Y_t = Y_0 + \int_0^t b_Y(s)ds + \int_0^t \sigma_Y(s)dW_Y(s) + \int_0^t \delta_Y \star (\mu - \nu)_t$$

Assume further that Y is square integrable. Assume $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i$, where each Z_i is a local Itô-martingale satisfying Definition 1:

$$Z_i(t) = \int_0^t \sigma_{Z_i}(s)dW_i(s) + \delta_{Z_i} \star (\mu_{Z_i} - \nu_{Z_i})_t$$

and each Z_i is square integrable. Assume that $[\bar{Z}_N, \bar{Z}_N]_T$ and $\langle \bar{Z}_N, \bar{Z}_N \rangle_T$ are bounded for all N. Divide the interval $[0,T]$ into M subintervals. Assume further that Y is either independent of Z_N or a square integrable martingale.

Then, it holds that

$$\sqrt{M}\left(\sum_{j=1}^{M}\Delta_j Y\Delta_j Z_N-[Y,Z_N]_T\right)=O_p(1)$$

Proof. **Step 1: Localization**
Using Theorem L.1 and following the same reasoning as in Section 4.4.1 of Jacod (2012), we can replace the local boundedness conditions with a bound on the whole time interval. I.e. without loss of generality, we can assume that there exists a constant C and a non-negative function Γ such that

$$\begin{aligned}
&\|\sigma_{Z_i}\|\leq C, && \|Z_i(t)\|\leq C, && \|\delta_{Z_i}\|^2\leq\Gamma, && \int\Gamma(z)\nu_{Z_i}(dz)\leq C\\
&\|\sigma_Y\|\leq C, && \|Y(t)\|\leq C, && \|\delta_Y\|^2\leq\Gamma, && \int\Gamma(z)\nu_Y(dz)\leq C\\
&\|b_Y\|\leq C
\end{aligned}$$

σ_{Z_N}, $\delta_{\bar{Z}_N}$ and $\nu_{\bar{Z}_N}$ are defined by

$$\langle\bar{Z}_N,\bar{Z}_N\rangle_t=\int_0^t\left(\sigma^2_{\bar{Z}_N}(s)+\int\delta^2_{\bar{Z}_N}(z,s)\nu_{\bar{Z}_N}(dz)\right)ds$$

Given our assumptions, we can use wlog that

$$\|\sigma_{\bar{Z}_N}\|\leq C,\qquad\|\bar{Z}_N(t)\|\leq C,\qquad\|\delta^2_{\bar{Z}_N}\|\leq\Gamma,\qquad\int\Gamma(z)\nu_{\bar{Z}_N}(dz)\leq C$$

Step 2: Bounds on increments
Denote the time increments by $\Delta_M=T/M$. Lemmas L.4, L.5 and L.6 together with the bounds on the characteristics of Y and Z_N imply that

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\leq s\leq\Delta_M}\|Y_{t+s}-Y_t\|^2\right]\leq & C\Delta_M\mathbb{E}\left[\int_t^{t+\Delta_M}\|b_Y(s)\|^2ds\right]+C\mathbb{E}\left[\int_t^{t+\Delta_M}\|\sigma_Y(s)\|^2ds\right]\\
&+C\mathbb{E}\left[\int_t^{t+\Delta_M}\int\|\delta_Y(s,z)\|^2\nu_Y(dz)ds\right]\leq\frac{C}{M}
\end{aligned}$$

and similarly

$$\mathbb{E}\left[\sup_{0\leq s\leq\Delta_M}\|\bar{Z}_N(s+t)-\bar{Z}_N(t)\|^2\right]\leq\frac{C}{M}$$

Step 3: Joint convergence

Define $G_{MN} = \sqrt{M}\left(\sum_{j=1}^{M} \Delta_j Y \Delta_j \bar{Z}_N - [Y, \bar{Z}_N]_T\right)$. We need to show, that $\forall \epsilon > 0$ there exists an n and a finite constant C such that

$$\mathbb{P}\left(\|G_{MN}\| > C\right) \leq \epsilon \qquad \forall M, N > n$$

By Markov' s inequality, if $\mathbb{E}\left[\|G_{MN}\|^2\right] < \infty$

$$\mathbb{P}\left(\|G_{MN}\| > C\right) \leq \frac{1}{C^2}\mathbb{E}\left[\|G_{MN}\|^2\right]$$

Hence it remains to show that $\mathbb{E}\left[\|G_{MN}\|^2\right] < \infty$ for $M, N \to \infty$.

Step 4: Bounds on sum of squared increments

By Itô's lemma, we have on each subinterval

$$\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, \bar{Z}_N] = \int_{t_j}^{t_{j+1}} (Y(s) - Y(t_j))d\bar{Z}_N(s) + \int_{t_j}^{t_{j+1}} (\bar{Z}_N(s) - \bar{Z}_N(t_j))dY(s)$$

As $\bar{Z}_N$ is square integrable and a local martingale, it is a martingale. By assumption Y is either independent of $\bar{Z}_N$ or a martingale as well. In the first case it holds that

$$\mathbb{E}\left[\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, \bar{Z}_N] | \mathfrak{F}_{t_j}\right] = \mathbb{E}\left[\Delta_j Y | \mathfrak{F}_{t_j}\right] \mathbb{E}\left[\Delta_j \bar{Z}_N | \mathfrak{F}_{t_j}\right] = 0$$

In the second case both stochastic integrals $\int_0^t Y(s)d\bar{Z}_N(s)$ and $\int_0^t \bar{Z}_N(s)dY(s)$ are martingales. Hence in either case, $\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, Z_N]$ forms a sequence of martingale differences and we can apply Burkholder's inequality for discrete time martingales (Lemma L.2):

$$\begin{aligned}
\mathbb{E}\left[\|G_{MN}\|^2\right] &\leq M \sum_{j=1}^{M} \mathbb{E}\left[\|\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, \bar{Z}_N]\|^2\right] \\
&\leq M \sum_{j=1}^{M} \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (Y(s) - Y(t_j))d\bar{Z}_N(s) + \int_{t_j}^{t_{j+1}} (\bar{Z}_N(s) - \bar{Z}_N(t_j))dY(s)\|^2\right] \\
&\leq M \sum_{j=1}^{M} \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (Y(s) - Y(t_j))d\bar{Z}_N(s)\|^2\right] + M \sum_{j=1}^{M} \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(s) - \bar{Z}_N(t_j))dY(s)\|^2\right]
\end{aligned}$$

It is sufficient to show that $\mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (Y(s) - Y(t_j))d\bar{Z}_N\|^2\right] = \frac{C}{M^2}$ and $\mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(s) - \bar{Z}_N(t_j))dY\|^2\right]$

$= \frac{C}{M^2}$. By Lemma L.3 and step 1 and 2:

$$
\begin{aligned}
E\left[\| \int_{t_j}^{t_{j+1}} (Y(t) - Y(t_j)) d\bar{Z}_N \|^2\right] &\leq \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \|Y(t) - Y(t_j)\|^2 d\langle \bar{Z}_N \rangle\right] \\
&\leq \mathbb{E}\left[\int_0^T \|Y(t) - Y(t_j)\|^2 \left(\sigma^2_{\bar{Z}_N}(t) + \int \delta^2_{\bar{Z}_N}(z,t)\nu_{\bar{Z}_N}(z)\right) dt\right] \\
&\leq C\mathbb{E}\left[\int_{t_j}^{t_{j+1}} \|Y(t) - Y(t_j)\|^2 dt\right] \\
&\leq C\mathbb{E}\left[\sup_{t_j \leq t \leq t_{j+1}} \|Y(t) - Y(t_j)\|^2\right] \frac{1}{M} \\
&\leq \frac{C}{M^2}.
\end{aligned}
$$

Similarly using Lemma L.4 for the drift of Y and L.3 for the martingale part, we can bound the second integral:

$$
\begin{aligned}
\mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j)) dY \|^2\right] &\leq \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j)) b_Y dt \|^2\right] \\
&+ \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - Z_N(t_j)) \left(\sigma_Y dW_Y + \delta_Y d(\mu - \nu)\right) dt \|^2\right] \\
&\leq \frac{1}{M} C\mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 \|b_Y(t)\|^2 dt\right] \\
&+ CE\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 \left(\|\sigma_Y(t)\|^2 + \int \|\delta_Y\|^2(z,t)\nu_Y(z)\right) dt\right] \\
&\leq \frac{1}{M} C\mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 dt\right] \\
&+ C\mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 (t) dt\right] \\
&\leq C\mathbb{E}\left[\sup_{t_j \leq t \leq t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2\right] \frac{1}{M} \\
&\leq \frac{C}{M^2}
\end{aligned}
$$

Putting things together, we obtain:

$$
\mathbb{E}\left[\|G_{MN}\|^2\right] \leq M \sum_{j=1}^{M} \frac{C}{M^2} \leq C
$$

which proves the statement. ☐

Lemma C.1. *Assumption 1 holds. Then*

$$\frac{1}{N} F e \Lambda = O_p\left(\frac{1}{\sqrt{MN}}\right)$$

Proof. Apply Proposition C.1 with $Y = F$ and $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \Lambda_k e_k$. ☐

Lemma C.2. *Assumption 1 holds. Then*

$$\frac{1}{N} \sum_{k=1}^{N} \left(\sum_{j=1}^{M} e_{ji} e_{jk} - [e_i, e_k] \right) \Lambda_k = O_p\left(\frac{1}{\sqrt{MN}}\right)$$

Proof. Apply Proposition C.1 with $Y = e_i$ and $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \Lambda_k e_k$. ☐

Lemma C.3. *Assume Assumption 1 holds. Then*

$$\frac{1}{N} \sum_{i=1}^{N} \Lambda_i e_i(T) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

Proof. By Burkholder's inequality in Lemma L.3 we can bound

$$\mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} \Lambda_i e_i(T)\right)^2\right] \leq \mathbb{E}\left[\frac{1}{N^2} \Lambda^\top \langle e, e \rangle \Lambda\right] \leq \frac{C}{N}$$

based on Assumption 1. ☐

Lemma C.4. *Assume Assumption 1 holds. Then*

$$\sum_{j=1}^{M} e_{ji} e_{jk} - [e_i, e_k]_T = O_p\left(\frac{1}{\sqrt{M}}\right)$$

Proof. Apply Theorem L.2. ☐

Proof of Lemma 1:

Proof. If e_i has independent increments it trivially satisfies weak serial dependence. The harder part is to show that the second and third condition imply weak cross-sectional dependence. We need to show

$$\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[e_{j,i}^2 e_{j,r}^2\right] = O\left(\frac{1}{\delta}\right)$$

Step 1: Decompose the residuals into their continuous and jump component respectively:

$$\begin{aligned}
&\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[\left(e_{j,i}^C + e_{j,i}^D\right)^2 \left(e_{j,r}^C + e_{j,r}^D\right)^2\right] \\
\leq &C \sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \Bigg(\mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^C}^2\right] + \mathbb{E}\left[{e_{j,i}^D}^2 {e_{j,r}^D}^2\right] + \mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^D}^2\right] \\
&+ \mathbb{E}\left[e_{j,i}^C e_{j,i}^D {e_{j,r}^C}^2\right] + \mathbb{E}\left[e_{j,i}^C e_{j,i}^D {e_{j,i}^D}^2\right] + \mathbb{E}\left[e_{j,i}^C e_{j,i}^D e_{j,r}^C e_{j,r}^D\right] \Bigg).
\end{aligned}$$

Step 2: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^C}^2\right] = O_p\left(\frac{1}{\delta}\right)$
This is a a consequence the Cauchy-Schwartz inequality and Burkholder's inequality in Lemma L.3:

$$\mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^C}^2\right] \leq C\mathbb{E}\left[{e_{j,i}^C}^4\right]^{1/2} \mathbb{E}\left[{e_{j,r}^C}^4\right]^{1/2} \leq \frac{C}{M^2}$$

Step 3: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^D}^2 {e_{j,r}^D}^2\right] = O_p\left(\frac{1}{\delta}\right)$

$$\begin{aligned}
&\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^D}^2 {e_{j,r}^D}^2\right] \leq \max_{j,r} |{e_{j,r}^D}^2| \cdot \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E}\left[{e_{j,i}^D}^2\right] \\
\leq &C \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E}\left[\Delta_j \langle e_i^D, e_i^D \rangle\right] \leq \frac{C}{N} \mathbb{E}\left[\sum_{i=1}^{N} \langle e_i^D, e_i^D \rangle\right] \leq O\left(\frac{1}{\delta}\right)
\end{aligned}$$

where we have used the second and third condition.
Step 4: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[e_{j,i}^C e_{j,i}^D e_{j,r}^C e_{j,r}^D\right] = O_p\left(\frac{1}{\delta}\right)$

$$\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[e_{j,i}^{C}e_{j,i}^{D}e_{j,r}^{C}e_{j,r}^{D}\right]\leq\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[\sum_{j=1}^{M}|e_{j,i}^{D}||e_{j,r}^{D}|\sup_{j,i,r}\left(|e_{j,i}^{C}||e_{j,r}^{C}|\right)\right]$$

$$\leq C\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[\left(\sum_{j=1}^{M}{e_{j,i}^{D}}^2\right)^{1/2}\left(\sum_{j=1}^{M}{e_{j,r}^{D}}^2\right)^{1/2}\sup_{j,i}({e_{j,i}^{C}}^2)\right]$$

$$\leq C\mathbb{E}\left[\sup_{j,i}({e_{j,i}^{C}}^2)\right]\leq\frac{C}{M}.$$

Step 5: The other moments can be treated similarly as in step 2 to 4. □

Proposition C.2. Consequence of weak dependence
Assume Assumption 1 holds. If additionally Assumption 5, i.e. weak serial dependence and weak cross-sectional dependence, holds then we have:

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}\sum_{l=1}^{M}e_{li}=O_p\left(\frac{1}{\delta}\right)$$

Proof. By the localization procedure in Theorem L.1, we can assume without loss of generality that there exists a constant C such that

$$\|b_F(t)\|\leq C\qquad\|\sigma_F(t)\|\leq C\qquad\|F(t)\|\leq C\qquad\|\delta_F(t,z)\|^2\leq\Gamma(z)\qquad\int\Gamma(z)v_F(dz)\leq C$$

$$\|\sigma_{e_i}(t)\|\leq C\qquad\|e_i(t)\|\leq C\qquad\|\delta_{e_i}(t,z)\|^2\leq\Gamma(z)\qquad\int\Gamma(z)v_{e_i}(dz)\leq C$$

We want to show

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}e_i(T)=O_p\left(\frac{1}{\delta}\right)$$

where $e_i(T)=\sum_{l=1}^{M}e_{li}$. I proceed in several steps: First, I define

$$\tilde{Z}=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(F_j e_{ji}e_i(T)-\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]\right)$$

with the notation $\mathbb{E}_j[.]=\mathbb{E}[.|\mathfrak{F}_{t_j}]$ as the conditional expectation and $b_j^F=\int_{t_j}^{t_{j+1}}b^F(s)ds$ as the increment of the drift term of F. The proof relies on the repeated use of different Burkholder inequalities, in particular that $b_j^F=O_p\left(\frac{1}{M}\right)$, $\Delta_j\langle e_i,e_i\rangle=O_p\left(\frac{1}{M}\right)$ and $\mathbb{E}[F_j^2]\leq\frac{C}{M}$.

Step 1: To show $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]=O_p\left(\frac{1}{\delta}\right)$

$$\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]\right|\leq \sup|\mathbb{E}_j[b_j^F]|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}|\mathbb{E}_j\left[\Delta_j\langle e_i,e_i\rangle\right]|\leq O_p\left(\frac{1}{M}\right)O_p(1)$$

Step 2: To show: $\tilde{Z}=O_p\left(\frac{1}{\delta}\right)$
Note that by the independence assumption between F and e, the summands in $\tilde{Z}$ follow a martingale difference sequence. Thus, by Burkholder's inequality for discrete time martingales:

$$\begin{aligned}\mathbb{E}\left[\tilde{Z}^2\right]\leq & C\mathbb{E}\left[\sum_{j=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\left(F_je_{ji}e_i(T)-\mathbb{E}_j[b_j^F\Delta_j\langle e_i,e_i\rangle]\right)\right)^2\right]\\ \leq & C\mathbb{E}\Bigg[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2e_{ji}e_{jr}e_i(T)e_r(T)+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\left(\mathbb{E}_j[b_j^F]^2\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]\right)\\ & -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\left(F_je_{ji}e_i(T)\mathbb{E}_j[b_j^F]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]+F_je_{jr}e_r(T)\mathbb{E}_j[b_j^F]\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\right)\Bigg]\end{aligned}$$

The first term can be written as

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2e_{ji}e_{jr}e_i(T)e_r(T)\right]\\ =&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2e_{ji}e_{jr}\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]+\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2e_{ji}^2e_{jr}^2\right]\end{aligned}$$

Under the assumption of weak serial dependence in Assumption 5 the first sum is bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}e_{jr}\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]\\
\leq & C\left(\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}[F_j^2]|\mathbb{E}[e_{ji}e_{jr}]|\left|\mathbb{E}\left[\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]\right|\right)\\
\leq & C\left(\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}[F_j^2]|\mathbb{E}[e_{ji}e_{jr}]|\left|\mathbb{E}\left[\sum_{l\neq j}e_{li}e_{lr}\right]\right|\right)\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}|\mathbb{E}[\Delta_j\langle e_i,e_r\rangle]|\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}|\mathbb{E}\left[\int_{t_j}^{t_{j+1}}G_{i,r}(s)ds\right]|\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{r=1}^{N}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}\sum_{i=1}^{N}|G_{i,r}(s)|ds\right]\\
\leq & C\frac{1}{MN}
\end{aligned}
$$

Next, we turn to the second sum of the first term:

$$
\begin{aligned}
&\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[F_j^2\right]\mathbb{E}\left[e_{ji}^2e_{jr}^2\right]\\
\leq & \frac{C}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[e_{ji}^2e_{jr}^2\right]\\
\leq & \frac{C}{M\delta}
\end{aligned}
$$

In the last line, we have used weak cross-sectional dependence in Assumption 5. The third term can be bounded as follows

$$
\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[\mathbb{E}_j[b_j^F]^2\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]\right]\leq\frac{C}{M^2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\sum_{j=1}^{M}\frac{C}{M^2}\leq\frac{C}{M^3}
$$

The final two terms can be treated the same way:

$$\begin{aligned}
&\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[F_j e_{ji}e_i(T)\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]\right]\\
\leq&\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[F_j\mathbb{E}_j[b_j^F]\right]\mathbb{E}\left[e_{ji}e_i(T)\mathbb{E}_j\left[\Delta_j\langle e_r,e_r\rangle\right]\right]\\
\leq&\sum_{j=1}^{M}\mathbb{E}\left[F_j\mathbb{E}_j[b_j^F]\right]\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}e_{ji}e_i(T)\right|\mathbb{E}_j\left[\frac{1}{N}\sum_{r=1}^{N}\Delta_j\langle e_r,e_r\rangle\right]\right]\\
\leq&\frac{C}{M^{3/2}}\sum_{j=1}^{M}\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}e_{ji}e_i(T)\right|\right]\frac{C}{M}\\
\leq&\frac{C}{M^{3/2}}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[|e_{ji}|\right]\leq\frac{C}{M^2}
\end{aligned}$$

□

Lemma C.5. Convergence rate of sum of residual increments: *Under Assumptions 1 and 2 it follows that*

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i e_{j,i}=O_p\left(\frac{1}{\delta}\right)$$

Proof. We apply Burkholder's inequality from Lemma L.3 together with Theorem L.1:

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i e_{j,i}\right)^2\right]\leq C\mathbb{E}\left[\frac{1}{N^2}\Lambda^\top\Delta_j\langle e,e\rangle\Lambda\right]\leq C\mathbb{E}\left[\frac{1}{N^2}\Lambda^\top\int_{t_j}^{t_{j+1}}G(s)ds\Lambda\right]\leq\frac{C}{NM}$$

which implies

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i e_{j,i}=O_p\left(\frac{1}{\sqrt{NM}}\right).$$

□

C.2 Central Limit Theorems

Lemma C.6. Central limit theorem for covariation between F and e_i

Assume that Assumptions 1 and 2 hold. Then

$$\sqrt{M}\sum_{j=1}^{M} F_j e_{ji} \stackrel{L-s}{\to} N(0,\Gamma_i)$$

where the entry $\{l,g\}$ of the $K\times K$ matrix Γ_i is given by

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g}\sigma^2_{e_i} ds + \sum_{s\le T}\Delta F^l(s)\Delta F^g(s)\sigma^2_{e_i}(s) + \sum_{s'\le T}\Delta e_i^2(s')\sigma_{F^g,F^l}(s')$$

F^l denotes the l-th component of the the K dimensional process F and σ_{F^l,F^g} are the entries of its $K\times K$ dimensional volatility matrix.

Proof. Apply Theorem L.2 using that independence of F and e_i implies $[F,e_i]=0$. □

Lemma C.7. Martingale central limit theorem with stable convergence to Gaussian martingale

Assume $Z^n(t)$ is a sequence of local square integrable martingales and Z is a Gaussian martingale with quadratic characteristic $\langle Z,Z\rangle$. Assume that for any $t>0$

1. $\int_0^t\int_{|z|>\epsilon} z^2\nu^n(ds,dx) \stackrel{p}{\to} 0 \ \forall\epsilon\in(0,1]$
2. $[Z^n,Z^n]_t \stackrel{p}{\to} [Z,Z]_t$

Then $Z^n \stackrel{L-s}{\to} Z$.

Proof. The convergence in distribution follows immediately from Lemma L.1. In order to show the stable weak convergence in Theorem L.4, we need to show that the nesting condition for the filtration holds. We construct a triangular array sequence $X^n(t) = Z^n([tk_n])$ for $0\le t\le 1$ and some $k_n\to\infty$. The sequence of histories is $\mathfrak{F}^n_t = \mathfrak{H}^n_{[tk_n]}; 0\le t\le 1$, where $\mathfrak{H}^n$ is the history of Z^n. Now, $t_n = \frac{1}{\sqrt{k_n}}$ is a sequence that satisfies the nesting condition. □

Lemma C.8. Martingale central limit theorem for sum or residuals

Assume that Assumption 1 is satisfied and hence, in particular $e_i(t)$ are square integrable martingales. Define $Z_N = \frac{1}{\sqrt{N}}\sum_{i=1}^N \Lambda_i e(t)$. Assume that for any $t>0$

1. $\frac{1}{N}\Lambda^\top\langle e,e\rangle^D_t\Lambda \stackrel{p}{\to} 0$
2. $\frac{1}{N}\Lambda^\top[e,e]^D_t\Lambda \stackrel{p}{\to} 0$
3. $\frac{1}{N}\Lambda^\top[e,e]_t\Lambda \stackrel{p}{\to} \Phi_t$

Then, conditioned on its quadratic covariation Z_N converges stably in law to a normal distribution.

$$Z_N \stackrel{L-s}{\to} N(0,\Phi_t).$$

Proof. By Lemma C.7 $Z_N \overset{L-s}{\to} Z$, where Z is a Gaussian process with $\langle Z, Z\rangle_t = \Phi_t$. Conditioned on its quadratic variation, the stochastic process evaluated at time t has a normal distribution. □

D Estimation of the Loadings

Lemma D.1. A decomposition of the loadings estimator
Let V_{MN} be the $K \times K$ matrix of the first K largest eigenvalues of $\frac{1}{N}X^\top X$. Define $H = \frac{1}{N}\left(F^\top F\right)\Lambda^\top \hat{\Lambda} V_{MN}^{-1}$. Then we have the decomposition

$$V_{MN}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = \frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k [e_i, e_k]_T + \frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \phi_{ki} + \frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \eta_{ki} + \frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \xi_{ki}$$

with

$$\phi_{ki} = \sum_{j=1}^M e_{ji}e_{jk} - [e_i, e_k]_T$$
$$\eta_{ki} = \Lambda_k^\top \sum_{j=1}^M F_j e_{ji}$$
$$\xi_{ki} = \Lambda_i^\top \sum_{j=1}^M F_j e_{jk}$$

Proof. This is essentially the identity in the proof of Theorem 1 in Bai and Ng (2002). From

$$\left(\frac{1}{N}X^\top X\right)\hat{\Lambda} = \hat{\Lambda} V_{MN}$$

it follws that $\frac{1}{N}X^\top X \hat{\Lambda} V_{MN}^{-1} = \hat{\Lambda}$. Substituting the definition of X, we obtain

$$\left(\hat{\Lambda} - \Lambda H\right) V_{MN} = \frac{1}{N}e^\top e\hat{\Lambda} + \frac{1}{N}\Lambda F^\top F \Lambda^\top \hat{\Lambda} + \frac{1}{N}e^\top F\Lambda^\top\hat{\Lambda} + \frac{1}{N}\Lambda F^\top e\hat{\Lambda} - \Lambda H V_{MN}$$

H is chosen to set

$$\frac{1}{N}\Lambda F^\top F\Lambda^\top\hat{\Lambda} - \Lambda H V_{MN} = 0.$$

□

Lemma D.2. Mean square convergence of loadings estimator *Assume Assumption 1* *holds. Then*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - H^\top \Lambda_i\|^2 = O_p\left(\frac{1}{\delta}\right).$$

Proof. This is essentially Theorem 1 in Bai and Ng (2002) reformulated for the quadratic variation and the proof is very similar. In Lemma D.4 it is shown that $\|V_{MN}\| = O_p(1)$. As $(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)$, we have $\|\hat{\Lambda}_i - \Lambda_i H\|^2 \leq (a_i + b_i + c_i + d_i)\cdot O_p(1)$ with

$$\begin{aligned}
a_i &= \frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k[e_k, e_i]\|^2 \\
b_i &= \frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki}\|^2 \\
c_i &= \frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}\|^2 \\
d_i &= \frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{kI}\|^2
\end{aligned}$$

Step 1: To show: $\frac{1}{N}\sum_{i=1}^{N} a_i = O_p\left(\frac{1}{N}\right)$

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} a_i &\leq \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k[e_k, e_i]\|^2\right) \\
&\leq \frac{1}{N}\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}[e_k, e_i]_T^2\right) \\
&= O_p\left(\frac{1}{N}\right)
\end{aligned}$$

The first term is $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_k\|^2 = O_p(1)$. The second term can be bounded by using the norm equivalence between the Frobenius and the spectral norm. Note that $\sum_{i=1}^{N}\sum_{k=1}^{N}[e_k, e_i]_T^2$ is simply the squared Frobenius norm of the matrix $[e, e]$. It is well-known that any $N \times N$ matrix A with rank N satisfies $\|A\|_F \leq \sqrt{N}\|A\|_2$. Therefore

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}[e_k, e_i]_T^2 \leq \|[e, e]\|_2^2 = O_p(1).$$

Step 2: To show: $\frac{1}{N}\sum_{i=1}^{N} b_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} b_i &\leq \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki}\|^2\right) \\
&\leq \frac{1}{N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\hat{\Lambda}_k^\top\hat{\Lambda}_l\phi_{ki}\phi_{li} \\
&\leq \frac{1}{N}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\hat{\Lambda}_k^\top\hat{\Lambda}_l\right)^2\right)^{1/2}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2\right)^{1/2} \\
&\leq \frac{1}{N}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\hat{\Lambda}_k^\top\hat{\Lambda}_l\right)^{1/2}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2\right)^{1/2}
\end{aligned}
$$

The second term is bounded by

$$
\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2 \leq N^2\max_{k,l}\phi_{kl}^4
$$

As $\phi_{kl}^4 = \left(\sum_{j=1}^{M} e_{jk}e_{jl} - [e_k, e_l]\right)^4 = O_p\left(\frac{1}{M^2}\right)$, we conclude

$$
\frac{1}{N}\sum_{i=1}^{N} b_i \leq \frac{1}{N}O_p\left(\frac{N}{M}\right) = O_p\left(\frac{1}{M}\right)
$$

Step 3: To show: $\frac{1}{N}\sum_{i=1}^{N} c_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N^3}\sum_{i=1}^{N}\|\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}\|^2 &\leq \frac{1}{N}\sum_{i=1}^{N}\|F^\top e_i\|^2\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k\|^2\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right) \\
&\leq \frac{1}{N}\left(\sum_{i=1}^{N}\|F^\top e_i\|^2\right)O_p(1) \leq O_p\left(\frac{1}{M}\right)
\end{aligned}
$$

The statement is a consequence of Lemma C.6.
Step 4: To show: $\frac{1}{N}\sum_{i=1}^{N} d_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}\|^2 &= \frac{1}{N^2}\|\sum_{k=1}^{N}\sum_{j=1}^{M}\hat{\Lambda}_k\Lambda_i^\top F_j e_{jk}\|^2 \\
&\leq \|\Lambda_i\|^2\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k\|^2\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\sum_{j=1}^{M}F_j e_{jk}\|^2\right)
\end{aligned}
$$

The statement follows again from Lemma C.6.

Step 5: From the previous four steps we conclude

$$\frac{1}{N}\sum_{i=1}^{N}(a_i+b_i+c_i+d_i)=O_p\left(\frac{1}{\delta}\right)$$

$\square$

Lemma D.3. Convergence rates for components of loadings estimator
Under Assumptions 1 and 2, it follows that

1. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_k,e_i]_T=O_p\left(\frac{1}{\sqrt{N\delta}}\right)$
2. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki}=O_p\left(\frac{1}{\sqrt{M\delta}}\right)$
3. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}=O_p\left(\frac{1}{\sqrt{\delta}}\right)$
4. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}=O_p\left(\frac{1}{\sqrt{M\delta}}\right)$

Proof. This is essentially Lemma A.2 in Bai (2003). The proof follows a similar logic to derive a set of inequalities. The main difference is that we use Lemmas C.1, C.2, C.4 and C.6 for determining the rates.
Proof of (1.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_k,e_i]=\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k-H^\top\Lambda_k\right)[e_k,e_i]+\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_k,e_i]$$

The second term can be bounded using Assumption 2

$$\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_k,e_i]\leq\max_k\|\Lambda_k\|\|H\|\frac{1}{N}\sum_{k=1}^{N}\|[e_k,e_i]\|=O_p\left(\frac{1}{N}\right)$$

For the first term we use Lemma D.2:

$$\begin{aligned}\left\|\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k-H^\top\Lambda_k\right)[e_k,e_i]\right\|&\leq\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k-H^\top\Lambda_k\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\sum_{k=1}^{N}[e_k,e_i]^2\right)^{1/2}\\&=O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)=O_p\left(\frac{1}{\sqrt{N\delta}}\right)\end{aligned}$$

The local boundedness of every entry of $[e,e]$ and Assumption 2 imply that

$$\sum_{k=1}^{N}\|[e_k,e_i]\|^2\leq max_{l=1,\dots N}\|[e_l,e_i]\|\sum_{k=1}^{N}\|[e_k,e_i]\|=O_p(1)$$

Proof of (2.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki} = \frac{1}{N}\sum_{k=1}^{N}\phi_{ki}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right) + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\phi_{ki}$$

Using Lemma C.4 we conclude that the first term is bounded by

$$\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\sum_{j=1}^{M}e_{ji}e_{jk} - [e_i,e_k]_T\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{M}}\right)$$

The second term is $O_p\left(\frac{1}{\sqrt{M\delta}}\right)$ by Lemma C.4.

Proof of (3.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki} = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\Lambda_k{}^\top F^\top e_i + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\Lambda_k{}^\top F^\top e_i$$

Applying the Cauchy-Schwartz inequality to the first term yields

$$\begin{aligned}
\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki} &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\eta_{ki}^2\right)^{1/2} \\
&\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\|F^\top e_i\|^2\right)^{1/2} \\
&\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\|F^\top e_i\|^2\right)^{1/2} \leq O_p\left(\frac{1}{\sqrt{\delta M}}\right).
\end{aligned}$$

For the second term we obtain the following bound based on Lemma C.6:

$$\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\Lambda_k{}^\top F^\top e_i = H^\top\left(\frac{1}{N}\sum_{k=1}^{N}\Lambda_k\Lambda_k{}^\top\right)\left(F^\top e_i\right) \leq O_p\left(\frac{1}{\sqrt{M}}\right)$$

Proof of (4.): We start with the familiar decomposition

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki} = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki} + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}$$

The first term is bounded by

$$\begin{aligned}
\|\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top \Lambda_k\right)\Lambda_i{}^\top F^\top e_k\| &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top \Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|F^\top e_k\|^2\right)^{1/2}\|\Lambda_i\| \\
&\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|F^\top e_k\|^2\right)^{1/2} \\
&\leq O_p\left(\frac{1}{\sqrt{\delta M}}\right)
\end{aligned}$$

The rate of the second term is a direct consequence of Proposition C.1:

$$\frac{1}{N}\sum_{k=1}^{N} H^\top \Lambda_k e_k^\top F \Lambda_i = O_p\left(\frac{1}{\sqrt{MN}}\right)$$

This very last step is also different from the Bai (2003) paper. They essentially impose this last conversion rate as an assumption (Assumption F.2), while I derive explicit conditions for the stochastic processes in Proposition C.1. □

Lemma D.4. Limit of V_{MN}
Assume Assumptions 1 and 2 hold. For $M, N \to \infty$, we have

$$\frac{1}{N}\hat{\Lambda}^\top\left(\frac{1}{N}X^\top X\right)\hat{\Lambda} = V_{MN} \xrightarrow{p} V$$

and

$$\frac{\hat{\Lambda}^\top \Lambda}{N}\left(F^\top F\right)\frac{\Lambda^\top \hat{\Lambda}}{N} \xrightarrow{p} V$$

where V is the diagonal matrix of the eigenvalues of $\Sigma_\Lambda^{1/2^\top}\Sigma_F \Sigma_\Lambda^{1/2}$

Proof. See Lemma A.3 in Bai (2003) and the paper by Stock and Watson (2002b). □

Lemma D.5. The matrix Q
Under Assumptions 1 and 2

$$plim_{M,N\to\infty}\frac{\hat{\Lambda}^\top \Lambda}{N} = Q$$

where the invertible matrix Q is given by $V^{1/2}\Upsilon^\top \Sigma_F^{-1/2}$ with Υ being the eigenvector of $\Sigma_F^{1/2}\Sigma_\Lambda \Sigma_F^{1/2}$

Proof. The statement is essentially Proposition 1 in Bai (2003) and the proof follows the same logic. Starting with the equality $\frac{1}{N}X^\top X\hat{\Lambda} = \hat{\Lambda}V_{MN}$, we multiply both sides by $\frac{1}{N}(F^\top F)^{1/2}\Lambda^\top$

to obtain

$$(F^\top F)^{1/2}\frac{1}{N}\Lambda^\top\left(\frac{X^\top X}{N}\right)\hat{\Lambda}=(F^\top F)^{1/2}\left(\frac{\Lambda^\top\hat{\Lambda}}{N}\right)V_{MN}$$

Plugging in $X=F\Lambda^\top+e$, we get

$$(F^\top F)^{1/2}\left(\frac{\Lambda^\top\Lambda}{N}\right)(F^\top F)\left(\frac{\Lambda^\top\hat{\Lambda}}{N}\right)+d_{NM}=(F^\top F)^{1/2}\left(\frac{\Lambda^\top\hat{\Lambda}}{N}\right)V_{MN}$$

with

$$d_{NM}=(F^\top F)^{1/2}\left(\frac{\Lambda^\top e^\top F}{N}\frac{\Lambda^\top\hat{\Lambda}}{N}+\frac{\Lambda^\top\Lambda}{N}\frac{F^\top e\hat{\Lambda}}{N}+\frac{\Lambda^\top e^\top e\hat{\Lambda}}{N^2}\right)$$

Applying Lemmas C.1 and C.2, we conclude $d_{NM}=o_p(1)$. The rest of the proof is essentially identical to Bai's proof. □

Lemma D.6. Properties of Q and H *Under Assumptions 1 and 2*

1. *plim*$_{M,N\to\infty}H=Q^{-1}$
2. $Q^\top Q=\Sigma_\Lambda$
3. *plim*$_{M,N\to\infty}HH^\top=\Sigma_\Lambda^{-1}$

Proof. Lemma D.5 yields $H=(F^\top F)\left(\frac{\Lambda^\top\hat{\Lambda}}{N}\right)V^{-1}\xrightarrow{p}\Sigma_F Q^\top V^{-1}$ and the definition of V is $\Upsilon V\Upsilon^\top={\Sigma_F^{1/2}}^\top\Sigma_\Lambda\Sigma_F^{1/2}$. Hence, the first statement follows from

$$\begin{aligned}H^\top Q&=V^{-1}Q\Sigma_F Q^\top+o_p(1)\\&=V^{-1}V^{1/2}\Upsilon^\top\Sigma_F^{-1/2}\Sigma_F{\Sigma_F^{-1/2}}^\top\Upsilon V^{1/2}+o_p(1)\\&=V^{-1}V+o_p(1)=I+o_p(1)\end{aligned}$$

The second statement follows from the definitions:

$$\begin{aligned}Q^\top Q&={\Sigma_F^{-1/2}}^\top\Upsilon V^{1/2}V^{1/2}\Upsilon^\top\Sigma_F^{1/2}\\&={\Sigma_F^{-1/2}}^\top{\Sigma_F^{1/2}}^\top\Sigma_\Lambda\Sigma_F^{1/2}\Sigma_F^{-1/2}\\&=\Sigma_\Lambda\end{aligned}$$

The third statement is a simple combination of the first two statements. □

Proof of Theorem 3:

Proof. Except for the asymptotic distribution of $\sqrt{M}F^\top e_i$, the proof is the same as for Theorem 1 in Bai (2003). By Lemma D.3

$$\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) V_{MN} = O_p\left(\frac{1}{\sqrt{M\delta}}\right) + O_p\left(\frac{1}{\sqrt{N\delta}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{M\delta}}\right)$$

The dominant term is $\frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \eta_{ki}$. Hence, we get the expansion

$$\sqrt{M}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = V_{MN}^{-1}\frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \Lambda_k^\top \sqrt{M} F^\top e_i + O_p\left(\frac{\sqrt{M}}{\delta}\right)$$

If $\frac{\sqrt{M}}{N} \to 0$, then using Lemmas C.6 and D.5, we obtain

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N\left(0, V^{-1} Q \Gamma_i Q^\top V^{-1}\right)$$

If $\liminf \frac{\sqrt{M}}{N} \geq \tau > 0$, then

$$N(\hat{\Lambda}_i - \Lambda_i H) = O_p\left(\frac{N}{\sqrt{M\delta}}\right) + O_p\left(\frac{\sqrt{N}}{\sqrt{\delta}}\right) + O_p\left(\frac{N}{\sqrt{M}}\right) + O_p\left(\frac{N}{\sqrt{M\delta}}\right) = O_p(1)$$

□

Lemma D.7. Consistency of loadings
Assume Assumption 1 holds. Then

$$\hat{\Lambda}_i - H^\top \Lambda_i = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

Proof. If we impose additionally Assumption 2, then this lemma is a trivial consequence of Theorem 3. However, even without Assumption 2, Lemma D.3 can be modified to show that

$$V_{MN}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = O_p\left(\frac{1}{\sqrt{\delta}}\right) + O_p\left(\frac{1}{\sqrt{N\delta}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{M\delta}}\right).$$

□

E Estimation of the Factors

Lemma E.1. *Assume that Assumptions 1 and 2 hold. Then*

$$\sum_{j=1}^{M} \frac{1}{N} F_j (\Lambda - \hat{\Lambda} H^{-1})^\top \hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$$

Proof. The overall logic of the proof is similar to Lemma B.1 in Bai (2003), but the underlying conditions and derivations of the final bounds are different. It is sufficient to show that

$$\frac{1}{N}(\hat{\Lambda} - \Lambda H)^\top \Lambda = O_p\left(\frac{1}{\delta}\right).$$

First using Lemma D.1 we decompose this term into

$$\begin{aligned}\frac{1}{N}(\hat{\Lambda} - \Lambda H)^\top \Lambda &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \phi_{ik} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k [e_i, e_k] + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \eta_{ki} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \xi_{ki}\right)\Lambda_i{}^\top \\ &= I + II + III + IV\end{aligned}$$

We will tackle all four terms one-by-one.
Term I: The first term can again be decomposed into

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k \phi_{ik} \Lambda_i{}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top \Lambda_k)\phi_{ik}\Lambda_i{}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \phi_{ik} \Lambda_i{}^\top$$

Due to Lemmas C.2 and D.2 the first term of I is bounded by

$$\begin{aligned}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top \Lambda_k)\phi_{ik}\Lambda_i{}^\top &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top \Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\phi_{ik}\Lambda_i{}^\top\|^2\right)^{1/2} \\ &\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}(e_{ji}e_{jk} - [e_i, e_k])\Lambda_i{}^\top\|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\sqrt{\delta}}\right) O_p\left(\frac{1}{\sqrt{MN}}\right)\end{aligned}$$

Now we turn to the second term, which we can bound using Lemma C.2 again:

$$\begin{aligned}
\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \phi_{ik} {\Lambda_i}^\top\| &\leq \|H\| \|\frac{1}{N}\sum_{k=1}^{N}\Lambda_k \frac{1}{N}\sum_{i=1}^{N}\phi_{ik}{\Lambda_i}^\top\| \\
&\leq O_p(1)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\phi_{ik}{\Lambda_i}^\top\|^2\right)^{1/2} \\
&\leq O_p\left(\frac{1}{\sqrt{MN}}\right)
\end{aligned}$$

Hence, I is bounded by the rate $O_p\left(\frac{1}{\sqrt{MN}}\right)$.

Term II: Next we deal with II:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_i,e_k]{\Lambda_i}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i,e_k]{\Lambda_i}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_i,e_k]{\Lambda_i}^\top$$

Cauchy-Schwartz applied to the first term yields

$$\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i,e_k]{\Lambda_i}^\top &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}[e_i,e_k]{\Lambda_i}^\top\|^2\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{\delta}N}\right)
\end{aligned}$$

We used Lemma D.2 for the first factor and Assumption 2 in addition with the boundedness of $\|\Lambda_i\|$ for the second factor. By the same argument the second term of II converges at the following rate

$$\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_i,e_k]{\Lambda_i}^\top &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}[e_i,e_k]{\Lambda_i}^\top\|^2\right)^{1/2} \\
&\leq O_p\left(\frac{1}{N}\right)
\end{aligned}$$

Thus, the rate of II is $O_p\left(\frac{1}{N}\right)$. Next, we address III.

Term III: We start with the familiar decomposition

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}{\Lambda_i}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki}{\Lambda_i}^\top + \frac{1}{N^2}\sum_{k=1}^{N}\sum_{i=1}^{N}H^\top\Lambda_k\eta_{ki}{\Lambda_i}^\top$$

We use Lemmas C.1 and D.2 and the boundedness of $\|\Lambda_k\|$. The first term is bounded by

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki}\Lambda_i{}^\top \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\Lambda_k{}^\top F_j e_{ji}\Lambda_i\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{\delta NM}}\right)$$

The second term is bounded by

$$\frac{1}{N^2}\sum_{k=1}^{N}\sum_{i=1}^{N}H^\top\Lambda_k\eta_{ki}\Lambda_i{}^\top \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\Lambda_k{}^\top F_j e_{ji}\Lambda_i\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{NM}}\right)$$

This implies that III is bounded by $O_p\left(\frac{1}{\sqrt{MN}}\right)$.
Term IV: Finally, we deal with IV:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}\Lambda_i{}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki}\Lambda_i{}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}\Lambda_i{}^\top.$$

The first term can be bounded using Lemmas D.2 and Lemma C.6:

$$\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki}\Lambda_i{}^\top\| \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\Lambda_i{}^\top F^\top e_i\Lambda_i{}^\top\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{\delta M}}\right)$$

For the second term we need the boundedness of Λ_i and a modification of Proposition C.1:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}\Lambda_i{}^\top\| &= \|\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{M}H^\top\Lambda_k e_{jk}F_j^\top\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i{}^\top\right)\| \\ &\leq \|\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i{}^\top\Lambda_i\right)\|\|\|\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{M}F_j e_{jk}\Lambda_k^\top H\| \\ &\leq O_p\left(\frac{1}{\sqrt{MN}}\right).\end{aligned}$$

In conclusion, IV is bounded by $O_p\left(\frac{1}{\sqrt{MN}}\right)$. Putting things together, we get

$$\frac{1}{N}(\hat{\Lambda}-\Lambda H)^\top\Lambda = O_p\left(\frac{1}{\sqrt{MN}}\right)+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{MN}}\right)+O_p\left(\frac{1}{\sqrt{MN}}\right)=O_p\left(\frac{1}{\delta}\right).$$

□

Lemma E.2. *Assume that Assumptions 1 and 2 hold. Then*

$$\sum_{j=1}^M\sum_{k=1}^N\frac{1}{N}\left(\hat{\Lambda}_k-H^\top\Lambda_k\right)e_{jk}=O_p\left(\frac{1}{\delta}\right)+O_p(1)\left(\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^M F_je_{ji}\sum_{l=1}^M e_{li}\right)$$

Without further assumptions the RHS is $O_p\left(\frac{1}{\delta}\right)+O_p\left(\frac{1}{\sqrt{M}}\right)$.

Proof. The general approach is similar to Lemma B.2 in Bai (2003), but the result is different, which has important implications for Theorem 4.

Note that $e_i(T)=\sum_{j=1}^M e_{ji}$. We want to show:

$$\frac{1}{N}\sum_{i=1}^N\left(\hat{\Lambda}_i-H^\top\Lambda_i\right)e_i(T)=O_p\left(\frac{1}{\delta}\right)+O_p(1)\left(\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^M F_je_{ji}\sum_{l=1}^M e_{li}\right).$$

We substitute the expression from Lemma D.1:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\left(\hat{\Lambda}_i-H^\top\Lambda_i\right)e_i(T)=&\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N\hat{\Lambda}_k[e_i,e_k]e_i(T)+\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N\hat{\Lambda}_k\phi_{ik}e_i(T)\\&+\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N\hat{\Lambda}_k\eta_{ik}e_i(T)+\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N\hat{\Lambda}_k\xi_{ik}e_i(T)\\&=I+II+III+IV\end{aligned}$$

Term I: We first decompose I into two parts:

$$\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N\hat{\Lambda}_k[e_i,e_k]e_i(T)=\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N(\hat{\Lambda}_k-H^\top\Lambda_k)[e_i,e_k]e_i(T)+\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N H^\top\Lambda_k[e_i,e_k]e_i(T).$$

Lemma D.2, Assumption 2 and the boundedness of $e_i(T)$ yield for the first term of I:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N(\hat{\Lambda}_k-H^\top\Lambda_k)[e_i,e_k]e_i(T)\|\le&\left(\frac{1}{N}\sum_{k=1}^N\|\hat{\Lambda}_k-H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^N\|\frac{1}{N}\sum_{i=1}^N e_i(T)[e_i,e_k]\|^2\right)^{1/2}\\\le&\, O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{N}\right).\end{aligned}$$

Using Assumption 2 , we bound the second term

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k [e_i, e_k] e_i(T) = O_p\left(\frac{1}{N}\right).$$

Hence, I is $O_p\left(\frac{1}{N}\right)$.

Term II: We split II into two parts:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} \hat{\Lambda}_k \phi_{ik} e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} (\hat{\Lambda}_k - H^\top \Lambda_k)\phi_{ik} e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \phi_{ik} e_i(T)$$

As before we apply the Cauchy-Schwartz inequality to the first term and then we use Lemma C.4:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\phi_{ik}e_i(T)\| \leq & \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2} \\ & \cdot\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}e_i(T)\left(\sum_{j=1}^{M}e_{ji}e_{jk} - [e_i, e_k]\right)\|^2\right)^{1/2} \\ \leq & O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{M}}\right)\end{aligned}$$

The second term can be bounded by using a modification of Lemma C.2 and the boundedness of $e_i(T)$:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \left(\sum_{j=1}^{M} e_{ji}e_{jk} - [e_i, e_k]\right) e_i(T) \leq O_p\left(\frac{1}{\sqrt{MN}}\right).$$

Thus, II is $O_p\left(\frac{1}{\sqrt{\delta M}}\right)$.

Term III: This term yields a convergence rate different from the rest and is responsible for the extra summand in the statement:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} \hat{\Lambda}_k \eta_{ik} e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ik} e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \eta_{ik} e_i(T)$$

The first term can be controlled using Lemma D.2 and Lemma C.6:

$$\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ik}e_i(T)\| \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}e_i(T)\Lambda_k{}^\top\sum_{j=1}^{M}F_je_{ji}\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{M}}\right)$$

Without further assumptions, the rate of the second term is slower than of all the other summands and can be calculated using Lemma C.6:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\Lambda_k{}^\top\sum_{j=1}^{M}F_je_{ji}e_i(T) = O_p(1)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_je_{ji}\sum_{l=1}^{M}e_{li}\right) = O_p\left(\frac{1}{\sqrt{M}}\right)$$

Term IV: We start with the usual decomposition for the last term:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ik}e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\xi_{ik}e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ik}e_i(T)$$

For the first term we use Lemma D.2 and Lemmas C.6 and C.8:

$$\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\xi_{ik}e_i(T)\| \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}e_i(T)\Lambda_i{}^\top\sum_{j=1}^{M}F_je_{jk}\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{\delta MN}}\right).$$

Similarly for the second term:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ik}e_i(T) = \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\left(\frac{1}{N}\sum_{i=1}^{N}e_i(T)\Lambda_i{}^\top\right)\left(\sum_{j=1}^{M}F_je_{jk}\right)$$
$$= O_p\left(\frac{1}{\sqrt{MN}}\right)$$

In conclusion, IV is $O_p\left(\frac{1}{\sqrt{MN}}\right)$. Putting the results together, we obtain

$$I + II + III + IV = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{\delta M}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{MN}}\right) = O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{M}}\right).$$

Term III is responsible for the low rate of convergence. □

Proof of Theorem 4:

Proof.

$$
\begin{aligned}
\hat{F} - F{H^{-1}}^\top &= \frac{1}{N} X\hat{\Lambda} - F{H^{-1}}^\top \\
&= (F(\Lambda - \hat{\Lambda}H^{-1} + \hat{\Lambda}H^{-1})^\top + e)\frac{1}{N}\hat{\Lambda} - F{H^{-1}}^\top \\
&= F\Lambda^\top \hat{\Lambda}\frac{1}{N} - F{H^{-1}}^\top \hat{\Lambda}^\top \hat{\Lambda}\frac{1}{N} + F{H^{-1}}^\top + e\hat{\Lambda}\frac{1}{N} - F{H^{-1}}^\top \\
&= \frac{1}{N} F(\Lambda - \hat{\Lambda}H^{-1})^\top \hat{\Lambda} + \frac{1}{N} e\hat{\Lambda} \\
&= \frac{1}{N} F(\Lambda - \hat{\Lambda}H^{-1})^\top \hat{\Lambda} + \frac{1}{N} e(\hat{\Lambda} - \Lambda H) + \frac{1}{N} e\Lambda H.
\end{aligned}
$$

By Lemmas E.1 and E.2, only the last term is of interest

$$
\begin{aligned}
\sum_{j=1}^{M} \left(\hat{F}_j - H^{-1}F_j\right) =& \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{N} \hat{\Lambda}_k \left(\Lambda_k - {H^{-1}}^\top \hat{\Lambda}_k\right)^\top F_j + \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{N} \left(\hat{\Lambda}_k - H^\top \Lambda_k\right) e_{jk} \\
&+ \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{N} H^\top \Lambda_k e_{jk} \\
=& O_p\left(\frac{1}{\delta}\right) + O_p(1) \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} F_j e_{ji} \sum_{l=1}^{M} e_{li}\right) + \frac{1}{N} e(T)\Lambda H.
\end{aligned}
$$

Under Assumption 5 Proposition C.2 implies $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji} \sum_{l=1}^{M} e_{li}\right) = O_p\left(\frac{1}{\delta}\right)$. If $\frac{\sqrt{N}}{M} \to 0$ then

$$
\sqrt{N} \sum_{j=1}^{M} \left(\hat{F}_j - H^{-1}F_j\right) = o_p(1) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H^\top \Lambda_i e_i(T)
$$

By Lemma C.8, we can apply the martingale central limit theorem and the desired result about the asymptotic mixed normality follows. In the case $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji} \sum_{l=1}^{M} e_{li}\right) = O_p\left(\frac{1}{\sqrt{M}}\right)$, the arguments are analogous. □

Lemma E.3. Consistency of factors
Assumptions 1 and 2 hold. Then $\hat{F}_T - H^{-1}F_T = O_p\left(\frac{1}{\sqrt{\delta}}\right)$.

Proof. The Burkholder-Davis-Gundy inequality in Lemma L.3 implies $\frac{1}{N} e_T \Lambda H = O_p\left(\frac{1}{\sqrt{N}}\right)$. In the proof of Theorem 4, we have shown that Assumptions 1 and 2 are sufficient for

$$
\sum_{j=1}^{M} \left(\hat{F}_j - H^{-1}F_j\right) = O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + \frac{1}{N} e_T \Lambda H.
$$

$\square$

Lemma E.4. Consistency of factor increments
Under Assumptions 1 and 2 we have

$$\hat{F}_j = H^{-1} F_j + O_p\left(\frac{1}{\delta}\right)$$

Proof. Using the same arguments as in the proof of Theorem 4 we obtain the decomposition

$$\hat{F}_j - H^{-1}F_j = \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\left(\Lambda_k - {H^{-1}}^\top\hat{\Lambda}_k\right)^\top F_j + \frac{1}{N}\sum_{k=1}^{N} e_{jk}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right) + \frac{1}{N}\sum_{k=1}^{N} H^\top\Lambda_k e_{jk}.$$

Lemma E.1 can easily be modified to show that

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\left(\Lambda_k - {H^{-1}}^\top\hat{\Lambda}_k\right)^\top F_j = O_p\left(\frac{1}{\delta}\right).$$

Lemma E.2 however requires some additional care. All the arguments go through for $e_{l,i}$ instead of $\sum_{l=1}^{M} e_{l,i}$ except for the term $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji} e_{li}\right)$. Based on our previous results we have $\sum_{j=1}^{M} F_j e_{j,i} = O_p\left(\frac{1}{\sqrt{M}}\right)$ and $e_{l,i} = O_p\left(\frac{1}{\sqrt{M}}\right)$. This yields

$$\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji} e_{li}\right) = O_p\left(\frac{1}{M}\right) = O_p\left(\frac{1}{\delta}\right)$$

Therefore

$$\frac{1}{N}\sum_{k=1}^{N} e_{jk}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right) = O_p\left(\frac{1}{\delta}\right).$$

Lemma C.5 provides the desired rate for the last term $\frac{1}{N}\sum_{k=1}^{N} H^\top\Lambda_k e_{jk} = O_p\left(\frac{1}{\delta}\right)$. $\square$

Lemma E.5. Consistent estimation of factor covariation
Under Assumptions 1 and 2 we can consistently estimate the quadratic covariation of the factors if $\frac{\sqrt{M}}{N} \to 0$. Assume $Y(t)$ is a stochastic process satisfying Definition 1. Then

$$\|\hat{F}^\top\hat{F} - H^{-1}[F,F]_T {H^{-1}}^\top\| = o_p(1) \qquad \|\sum_{j=1}^{M}\hat{F}_j Y_j - H^{-1}[F,Y]\| = o_p(1)$$

Proof. This is a simple application of Lemma E.4:

$$\begin{aligned}\sum_{j=1}^{M} \hat{F}_j \hat{F}_j^\top &= H^{-1} \left(\sum_{j=1}^{M} F_j F_j^\top \right) H^{-1^\top} + O_p\left(\frac{1}{\delta}\right) \sum_{j=1}^{M} |F_j| + \sum_{j=1}^{M} O_p\left(\frac{1}{\delta^2}\right) \\ &= H^{-1} \left(\sum_{j=1}^{M} F_j F_j^\top \right) H^{-1^\top} + O_p\left(\frac{\sqrt{M}}{\delta}\right) + O_p\left(\frac{M}{\delta^2}\right)\end{aligned}$$

By Theorem L.2

$$\left(\sum_{j=1}^{M} F_j F_j^\top \right) - [F, F]_T = O_p\left(\frac{1}{\sqrt{\delta}}\right)$$

The desired result follows for $\frac{\sqrt{M}}{N} \to 0$. The proof for $[F, Y]$ is analogous. □

F Estimation of Common Components

Proof of Theorem 5:

Proof. The proof is very similar to Theorem 3 in Bai (2003). For completeness I present it here:

$$\hat{C}_{T,i} - C_{T,i} = \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1} F_T + \hat{\Lambda}_i^\top \left(\hat{F}_T - H^{-1} F_T\right).$$

From Theorems 3 and 4 we have

$$\begin{aligned}\sqrt{\delta}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) &= \sqrt{\frac{\delta}{M}} V_{MN}^{-1} \frac{1}{N} \sum_{k=1}^{N} \hat{\Lambda}_k \Lambda_k^\top \sqrt{M} F^\top e_i + O_p\left(\frac{1}{\sqrt{\delta}}\right) \\ \sqrt{\delta}\left(\hat{F}_T - H^{-1} F_t\right) &= \sqrt{\frac{\delta}{M}} \sum_{i=1}^{N} H^\top \Lambda_i e_{T,i} + O_p\left(\sqrt{\frac{\delta}{M}}\right) + O_p\left(\frac{1}{\sqrt{\delta}}\right).\end{aligned}$$

If Assumption 5 holds, the last equation changes to

$$\sqrt{\delta}\left(\hat{F}_T - H^{-1} F_t\right) = \sqrt{\frac{\delta}{M}} \sum_{i=1}^{N} H^\top \Lambda_i e_{T,i} + O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

In the following, we will assume that weak serial dependence and cross-sectional dependence holds. The modification to the case without it is obvious. Putting the limit theorems for the loadings and the factors together yields:

$$\hat{C}_{T,i} - C_{T,i} = \sqrt{\frac{\delta}{M}} F^\top H^{-1^\top} V_{MN}^{-1} \left(\frac{1}{N} \Lambda^\top \hat{\Lambda} \right) \sqrt{M} F^\top e_i$$
$$+ \sqrt{\frac{\delta}{N}} \Lambda_i^\top H H^\top \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i e^{T,i} \right) + O_p \left(\frac{1}{\sqrt{\delta}} \right).$$

We have used

$$\hat{\Lambda}_i^\top \left(\hat{F}_T - H^{-1} F_T \right) = \Lambda_i^\top H (\hat{F}_T - H^{-1} F_T) + \left(\hat{\Lambda}_i^\top - \Lambda_i^\top H \right) \left(\hat{F}_T - H^{-1} F_T \right)$$
$$= \Lambda_i^\top H (\hat{F}_T - H^{-1} F_T) + O_p \left(\frac{1}{\delta} \right).$$

By the definition of H it holds that

$$H^{-1^\top} V_{MN}^{-1} \left(\frac{\hat{\Lambda}^\top \Lambda}{N} \right) = \left(F^\top F \right)^{-1}.$$

Using the reasoning behind Lemma D.6, it can easily be shown that

$$H H^\top = \left(\frac{1}{N} \Lambda^\top \Lambda \right)^{-1} + O_p \left(\frac{1}{\delta} \right).$$

Define

$$\xi_{NM} = F_T^\top \left(F^\top F \right)^{-1} \sqrt{M} F^\top e_i$$
$$\phi_{NM} = \Lambda_i^\top \left(\frac{1}{N} \Lambda^\top \Lambda \right)^{-1} \frac{1}{\sqrt{N}} \Lambda^\top e_T$$

By Lemmas C.6 and C.8, we know that these terms converge stably in law to a conditional normal distribution:

$$\xi_{NM} \stackrel{L-s}{\to} N(0, V_{T,i}) \qquad , \qquad \phi_{NM} \stackrel{L-s}{\to} N(0, W_{T,i})$$

Therefore,

$$\sqrt{\delta} \left(\hat{C}_{T,i} - C_{T,i} \right) = \sqrt{\frac{\delta}{M}} \xi_{NM} + \sqrt{\frac{\delta}{N}} \phi_{NM} + O_p \left(\frac{1}{\sqrt{\delta}} \right)$$

ξ_{NM} and ϕ_{NM} are asymptotically independent, because one is the sum of cross-sectional random variables, while the other is the sum of a particular time series of increments. If $\frac{\delta}{M}$ and $\frac{\delta}{N}$ converge, then asymptotic normality follows immediately from Slutzky's theorem. $\frac{\delta}{M}$ and $\frac{\delta}{N}$

are not restricted to be convergent sequences. We can apply an almost sure representation theory argument on the extension of the probability space similar to Bai (2003). □

Lemma F.1. Consistency of increments of common component estimator
Under Assumptions 1 and 2 it follows that

$$\hat{C}_{j,i} = C_{j,i} + O_p\left(\frac{1}{\delta}\right)$$
$$\hat{e}_{j,i} = e_{j,i} + O_p\left(\frac{1}{\delta}\right)$$

with $\hat{e}_{j,i} = X_{j,i} - \hat{C}_{j,i}$.

Proof. As in the proof for Theorem 5 we can separate the error into a component due to the loading estimation and one due to the factor estimation.

$$\hat{C}_{j,i} - C_{j,i} = \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1} F_j + \hat{\Lambda}_i^\top \left(\hat{F}_j - H^{-1} F_j\right).$$

By Lemmas D.7 and E.4 we can bound the error by $O_p\left(\frac{1}{\delta}\right)$. □

Lemma F.2. Consistent estimation of residual covariation *Assume Assumptions 1 and 2 hold. Then if* $\frac{\sqrt{M}}{\delta} \to 0$ *we have for* $i, k = 1, ..., N$ *and any stochastic process* $Y(t)$ *satisfying Definition 1:*

$$\sum_{j=1}^{M} \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1), \qquad \sum_{j=1}^{M} \hat{C}_{j,i}\hat{C}_{j,k} = [C_i, C_k] + o_p(1).$$
$$\sum_{j=1}^{M} \hat{e}_{j,i}Y_j = [e_i, Y] + o_p(1), \qquad \sum_{j=1}^{M} \hat{C}_{j,i}Y_j = [C_i, Y] + o_p(1).$$

Proof. Using Lemma F.1 we obtain

$$\sum_{j=1}^{M} \hat{e}_{j,i}\hat{e}_{j,k} = \sum_{j=1}^{M} e_{j,i}e_{j,k} + \sum_{j=1}^{M} O_p\left(\frac{1}{\delta^2}\right) + \sum_{j=1}^{M} |e_{j,i}| O_p\left(\frac{1}{\delta}\right) = \sum_{j=1}^{M} e_{j,i}e_{j,k} + o_p(1) = [e_i, e_k] + o_p(1).$$

The rest of the proof follows the same logic. □

Proof of Theorem 1:

Proof. This is a collection of the results in Lemmas D.7, E.3, E.5, F.1 and F.2. □

G Estimating Covariance Matrices

Proposition G.1. Consistent unfeasible estimator of covariance matrix of loadings *Assume Assumptions 1, 2 and 3 hold and* $\frac{\sqrt{M}}{N} \to 0$*. By Theorem 1*

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N\left(0, \Theta_\Lambda\right)$$

with

$$\Theta_{\Lambda,i} = V^{-1} Q \Gamma_i Q^\top V^{-1}$$

where the entry $\{l,g\}$ *of the* $K \times K$ *matrix* Γ_i *is given by*

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g} \sigma_{e_i}^2 ds + \sum_{s \leq T} \Delta F^l(s) \Delta F^g(s) \sigma_{e_i}^2(s) + \sum_{s' \leq T} \Delta e_i^2(s') \sigma_{F^g,F^l}(s').$$

F^l *denotes the* l*-th component of the the* K *dimensional process* F *and* σ_{F^l,F^g} *are the entries of its* $K \times K$ *dimensional volatility matrix. Take any sequence of integers* $k \to \infty$*,* $\frac{k}{M} \to 0$*. Denote by* $I(j)$ *a local window of length* $\frac{2k}{M}$ *around* j *with some* $\alpha > 0$ *and* $\omega \in \left(0, \frac{1}{2}\right)$*.*

Define a consistent, but unfeasible, estimator for Γ_i *by*

$$\begin{aligned}\bar{\Gamma}_{i,l,g} =& M \sum_{j=1}^{M} F_j^l F_j^g e_{j,i}^2 \mathbb{1}_{\{|F_j^l| \leq \alpha \Delta_M^\omega, |F_j^g| \leq \alpha \Delta_M^\omega, |e_{j,i}| \leq \alpha \Delta_M^\omega\}} \\ &+ \frac{M}{2k} \sum_{j=k+1}^{M-k} F_j^l F_j^g \mathbb{1}_{\{|F_j^l| \geq \alpha \Delta_M^\omega, |F_j^g| \geq \alpha \Delta_M^\omega\}} \left(\sum_{h \in I(j)} e_{h,i}^2 \mathbb{1}_{\{|e_{h,i}| \leq \alpha \Delta_M^\omega\}} \right) \\ &+ \frac{M}{2k} \sum_{j=k+1}^{M-k} e_{j,i}^2 \mathbb{1}_{\{|e_{j,i}| \geq \alpha \Delta_M^\omega\}} \left(\sum_{h \in I(j)} F_h^l F_h^g \mathbb{1}_{\{|F_h^l| \leq \alpha \Delta_M^\omega, |F_h^g| \leq \alpha \Delta_M^\omega\}} \right)\end{aligned}$$

Then

$$V_{MN}^{-1} \left(\frac{\hat{\Lambda}^\top \Lambda}{N} \right) \bar{\Gamma}_i \left(\frac{\Lambda^\top \hat{\Lambda}}{N} \right) V_{MN}^{-1} \xrightarrow{p} \Theta_{\Lambda,i}$$

Proof. The Estimator for Γ_i is an application of Theorem L.3. Note that we could generalize the statement to include infinite activity jumps as long as their activity index is smaller than 1. Finite activity jumps trivially satisfy this condition. The rest follows from Lemmas D.4 and D.5. □

Proof of Theorem 6:

Proof. By abuse of notation the matrix $e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ has elements $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ and the matrix $F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^\top$ has elements $F_j\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda_i^\top$. A similar notation is applied for other combinations of vectors with a truncation indicator function.

Step 1: To show: $\frac{1}{N}\hat{X}_j^C\hat{\Lambda} - \sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j = O_p\left(\frac{1}{\delta}\right)$

We start with a similar decomposition as in Theorem 4:

$$\begin{aligned}\frac{\hat{X}^C\hat{\Lambda}}{N} - F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}H^{-1\top}\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} =& \frac{1}{N}F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right)\\ &+ \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H.\end{aligned}$$

It can be shown that

$$\begin{aligned}\frac{1}{N}F_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} &= O_p\left(\frac{1}{\delta}\right)\\ \frac{1}{N}e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) &= O_p\left(\frac{1}{\delta}\right)\\ \frac{1}{N}e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H = \frac{1}{N}e_j^C\Lambda H + \frac{1}{N}\left(e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} - e_j^C\right)\Lambda H &= O_p\left(\frac{1}{\delta}\right).\end{aligned}$$

The first statement follows from Lemma E.1. The second one can be shown as in Lemma E.4. The first term of the third statement can be bounded using Lemma C.5. The rate for the second term of the third equality follows from the fact that the difference $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}} - e_{j,i}^C$ is equal to some drift term which is of order $O_p\left(\frac{1}{M}\right)$ and to $-\frac{1}{N}e_{j,i}^C$ if there is a jump in $X_{j,i}$.

Step 2: To show: $\frac{1}{N}\hat{X}_j^D\hat{\Lambda} - \sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j = O_p\left(\frac{1}{\delta}\right)$

As in step 1 we start with a decomposition

$$\begin{aligned}\frac{\hat{X}^D\hat{\Lambda}}{N} - F\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}H^{-1\top}\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} =& \frac{1}{N}F\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} + \frac{1}{N}e\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right)\\ &+ \frac{1}{N}e\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H.\end{aligned}$$

It follows

$$\begin{aligned}\frac{1}{N}F_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} &= O_p\left(\frac{1}{\delta}\right)\\ \frac{1}{N}e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) &= O_p\left(\frac{1}{\delta}\right)\\ \frac{1}{N}e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H = \frac{1}{N}e_j^D\Lambda H + \frac{1}{N}\left(e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}} - e_j^D\right)\Lambda H &= O_p\left(\frac{1}{\delta}\right).\end{aligned}$$

The first rate is a consequence of Lemma E.1, the second rate follows from Lemma D.7 and the third rate can be derived using similar arguments as in step 1.

Step 3: To show: $\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}} = O_p\left(\frac{1}{\delta}\right)$
By a similar decomposition as in Lemma F.1 we obtain

$$\begin{aligned}\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}} =& \left(\hat{\Lambda}_i - H^\top\Lambda_i\right)^\top H^{-1}F_j\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \hat{\Lambda}_i^\top\left(\frac{\hat{\Lambda}^\top \hat{X}^{C\top}_j}{N} - H^{-1}F_j\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right) \\ =&O_p\left(\frac{1}{\sqrt{\delta}}\right)\|F_j\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\| + O_p\left(\frac{1}{\delta}\right) \\ =&O_p\left(\frac{1}{\sqrt{\delta M}}\right) + O_p\left(\frac{1}{\delta}\right)\end{aligned}$$

The first rate follows from Lemma D.7 and the second rate can be deduced from step 1.
Step 4: To show $\hat{X}^D_{j,i} - \frac{\hat{X}^D_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}} = O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{\delta}}\right)\|F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\|$
A similar decomposition as in the previous step yields

$$\begin{aligned}\hat{X}^D_{j,i} - \frac{\hat{X}^D_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}} =& \left(\hat{\Lambda}_i - H^\top\Lambda_i\right)^\top H^{-1}F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \hat{\Lambda}_i^\top\left(\frac{\hat{\Lambda}^\top \hat{X}^{D\top}_j}{N} - H^{-1}F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\right) \\ \leq& O_p\left(\frac{1}{\sqrt{\delta}}\right)\|F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\| + O_p\left(\frac{1}{\delta}\right)\end{aligned}$$

where the first rate follows from Lemma D.7 and the second from step 2.
Step 5: To show: $M\sum_{j=1}^M\left(\frac{\hat{X}^C_j\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}^C_j\hat{\Lambda}}{N}\right)^\top\left(\hat{X}^C_{j,i} - \frac{\hat{X}^C_j\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2$
$= \mathrm{M}\sum_{j=1}^M\left(H^{-1}F_j\mathbb{1}_{\{|F_j|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)^\top\left(H^{-1}F_j\mathbb{1}_{\{|F_j|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\left(e^2_{j,i}\mathbb{1}_{\{|e_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right) + o_p(1)$
Step 1 and 3 yield

$$\begin{aligned}&M\sum_{j=1}^M\left(\frac{\hat{X}^C_j\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}^C_j\hat{\Lambda}}{N}\right)^\top\left(\hat{X}^C_{j,i} - \frac{\hat{X}^C_j\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2 \\ =&M\sum_{j=1}^M\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j\right)^\top\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j\right)\left(e^2_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right) + o_p(1)\end{aligned}$$

We need to show

$$\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j - H^{-1}F_j\mathbb{1}_{\{\{|F_j|\leq\alpha\Delta_M^{\bar{\omega}}\}} = o_p\left(\frac{1}{\sqrt{\delta}}\right).$$

By Mancini (2009) the threshold estimator correctly identifies the jumps for sufficiently large M. By Assumption 3 a jump in $X_{j,i}$ is equivalent to a jump in $\Lambda_i^\top F_j$ or/and a jump in $e_{j,i}$. Hence, it is sufficient to show that

$$\sum_{i=1}^N \mathbb{1}_{\{F_j^D \Lambda_i=0, e_i^D=0, |F_j^D|\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} + \sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} - I_K \sum_{i=1}^N \mathbb{1}_{\{e_{j,i}^D\neq 0, |F_j^D|=0\}} = o_p(1)$$

Note that

$$\begin{aligned}
\mathbb{P}\left(e_{j,i}^D \neq 0\right) &= \mathbb{E}\left[\mathbb{1}_{\{e_{j,i}^D\neq 0\}}\right] = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \int_{\mathbb{R}_{-0}} d\mu_{e_i}(ds,dx)\right] \\
&= \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \int_{\mathbb{R}_{-0}} d\nu_{e_i}(ds,dx)\right] \leq C \int_{t_j}^{t_{j+1}} ds = O\left(\frac{1}{M}\right).
\end{aligned}$$

It follows that $\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} = o_p(1)$ as

$$\mathbb{E}\left[\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\right] = \sum_{i=1}^N \mathbb{P}\left(e_i^D \neq 0\right) \frac{\hat{\Lambda}_i\hat{\Lambda}_i}{N} = O_p\left(\frac{1}{M}\right)$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\right)^2\right] &= \mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \hat{\Lambda}_i\hat{\Lambda}_i^\top\hat{\Lambda}_k\hat{\Lambda}_k^\top \mathbb{1}_{\{e_i^D\neq 0\}}\mathbb{1}_{\{e_k^D\neq 0\}}\right] \\
&\leq \left(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \|\hat{\Lambda}_i\hat{\Lambda}_i^\top\hat{\Lambda}_k\hat{\Lambda}_k^\top\|^2\right]\right)^{1/2} \left(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \mathbb{1}^2_{\{e_i^D\neq 0\}}\mathbb{1}^2_{\{e_k^D\neq 0\}}\right]\right)^{1/2} \\
&\leq C\left(\mathbb{E}\left[\sum_{t_j}^{t_{j+1}} \frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N G_{i,k} dt\right]\right)^{1/2} \leq \frac{C}{\sqrt{NM}}
\end{aligned}$$

By the same logic it follows that $\sum_{i=1}^N \mathbb{1}_{\{e_{j,i}^D\neq 0, |F_j^D|=0\}} = o_p(1)$. Last but not least

$$\begin{aligned}
\|\sum_{i=1}^N \mathbb{1}_{\{F_j^D\Lambda_i=0, e_i^D=0, |F_j^D|\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| &\leq \|\sum_{i=1}^N \mathbb{1}_{\{|F_j^D|\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| \\
&\leq \mathbb{1}_{\{|F_j^D|\neq 0\}} \|\sum_{i=1}^N \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| \leq O_p\left(\frac{1}{\sqrt{M}}\right).
\end{aligned}$$

On the other hand there are only finitely many j for which $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \neq e_{j,i}\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ and the difference is $O_p\left(\frac{1}{\sqrt{M}}\right)$, which does not matter asymptotically for calculating the multipower variation.

Step 6: To show: $\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\frac{\hat{X}_j^D\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_j^D\hat{\Lambda}}{N}\right)^\top\left(\sum_{h\in I(j)}\left(\hat{X}_{h,i}^C-\frac{\hat{X}_h^C\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2\right)$
$=\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar{\omega}}\}}\right)^\top\left(H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar{\omega}}\}}\right)\left(\sum_{h\in I(j)}\left(e_{h,i}^2\mathbb{1}_{\{|e_{h,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\right)+o_p(1)$
We start by plugging in our results from Steps 2 and 3:

$$\begin{aligned}
&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\frac{\hat{X}_j^D\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_j^D\hat{\Lambda}}{N}\right)^\top\left(\sum_{h\in I(j)}\left(\hat{X}_{h,i}^C-\frac{\hat{X}_h^C\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2\right)\\
=&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j\right)^\top\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j\right)\\
&\cdot\left(\sum_{h\in I(j)}\left(e_{h,i}^2\mathbb{1}_{\{|X_{h,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\right)+o_p(1).
\end{aligned}$$

We need to show that $\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j=H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar{\omega}}\}}+o_p\left(\frac{1}{\sqrt{\delta}}\right)$. This follows from

$$\sum_{i=1}^N\left(\mathbb{1}_{\{|F_j^D\Lambda_i|>0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}-I_K\mathbb{1}_{\{|F_j^D|\neq0\}}\right)-\sum_{i=1}^N\mathbb{1}_{\{|F_j^D\Lambda_i|>0,|F_j^D|>0,e_{j,i}^D=0\}}I_K+\sum_{i=1}^N\mathbb{1}_{\{e_{j,i}^D\neq0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}=o_p(1)$$

which can be shown by the same logic as in step 5.

Step 7: To show: $\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\hat{X}_{j,i}^D-\frac{\hat{X}_j^D\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2\left(\sum_{h\in I(j)}\left(\frac{\hat{X}_h^C\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_h^C\hat{\Lambda}}{N}\right)^\top\right)$
$=\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(e_{j,i}^2\mathbb{1}_{\{|e_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\right)\left(\sum_{h\in I(j)}\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)^\top\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\right)+o_p(1)$
In light of the previous steps we only need to show how to deal with the first term. By step 4 we have

$$\begin{aligned}
&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\hat{X}_{j,i}^D-\frac{\hat{X}_j^D\hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2\left(\sum_{h\in I(j)}\left(\frac{\hat{X}_h^C\hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_h^C\hat{\Lambda}}{N}\right)^\top\right)\\
=&\frac{M}{2k}\sum_{j\in J}\left(e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}+O_p\left(\frac{1}{\delta}\right)+O_P\left(\frac{1}{\sqrt{\delta}}\right)\|F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\|\right)^2\\
&\cdot\left(\sum_{h\in I(j)}\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)^\top\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\right)+o_p(1)
\end{aligned}$$

where J denotes the set of jumps of the process $X_i(t)$. Note that J contains only finitely many elements. The difference between $e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}$ and $e_{j,i}\mathbb{1}_{\{|e_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}$ is of order $O_p\left(\frac{1}{\sqrt{M}}\right)$ as there might be increments j where there is a jump in the factors but not in the residuals. As we consider only finitely many increments j the result follows. $\square$

Proof of Theorem 7:

Proof. Under cross-sectional independence of the error terms the asymptotic variance equals

$$\Theta_F = \underset{N,M\to\infty}{plim} H^\top \frac{\sum_{i=1}^N \Lambda_i [e_i, e_i] \Lambda_i^\top}{N} H$$

By Lemmas D.7 and F.2 we know that $\sum_{j=1}^M \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1)$ and $\hat{\Lambda}_i = H^\top \Lambda_i + O_p\left(\frac{1}{\sqrt{\delta}}\right)$ and the result follows immediately. $\square$

H Separating Continuous and Jump Factors

Lemma H.1. Convergence rates for truncated covariations
Under Assumptions 1 and 3 and for some $\alpha > 0$ *and* $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ *it follows that*

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M F_j e_{j,i} \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}\| = O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{N}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M F_j e_{j,i} \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta_M^{\bar{\omega}}\}}\| = O_p\left(\frac{1}{\sqrt{M}}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M \left(e_{j,i}e_{j,k} \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} \mathbb{1}_{\{|X_{j,k}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - [e_i^C, e_k^C]\right)\| = O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{N}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M \left(e_{j,i}e_{j,k} \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta_M^{\bar{\omega}}\}} \mathbb{1}_{\{|X_{j,k}| > \alpha \Delta_M^{\bar{\omega}}\}} - [e_i^D, e_k^D]\right)\| = O_p\left(\frac{1}{\sqrt{M}}\right).$$

Proof. I will only prove the first statement as the other three statements can be shown analogously. By Theorem L.6

$$\sum_{j=1}^M F_j e_{j,i} \mathbb{1}_{\{\|F_j\| \leq \alpha \Delta_M^{\bar{\omega}}, |e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} = O_p\left(\frac{1}{\sqrt{M}}\right).$$

However, as F and e_i are not observed our truncation is based on X. Hence we need to characterize

$$\sum_{j=1}^M F_j e_{j,i} \left(\mathbb{1}_{\{\|F_j\| \leq \alpha \Delta_M^{\bar{\omega}}, |e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}\right).$$

If there is a jump in X, there has to be also a jump in e_i or F. By Assumption 3 if there is a jump in e_i or $\Lambda_i^\top F$, there has to be a jump in X. However, it is possible that two factors F_k and F_l jump at the same time but their weighted average $\Lambda_i^\top F$ is equal to zero. Hence, we could not identify these jumps by observing only X_i. This can only happen for a finite number of indices i as $\lim_{N\to\infty} \frac{\Lambda^\top \Lambda}{N} = \Sigma_\Lambda$ has full rank. Hence

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{j=1}^{M} F_j e_{j,i}\left(\mathbb{1}_{\{\|F_j\|\leq \alpha\Delta_M^{\bar{\omega}},\ e_{j,i}\leq \alpha\Delta_M^{\bar{\omega}}\}} - \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)\right\| = O_p\left(\frac{1}{N}\right).$$

In the reverse case where we want to consider only the jump part, $|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}$ implies that either $\Lambda_i^\top F_j$ or $e_{j,i}$ has jumped. If we wrongly classify an increment $e_{j,i}$ as a jump although the jump happened in $\Lambda_i^\top F_j$, it has an asymptotically vanishing effect as we have only a finite number of jumps in total and the increment of a continuous process goes to zero with the rate $O_p\left(\frac{1}{\sqrt{M}}\right)$. □

Proof of Theorem 2:

Proof. I only prove the statement for the continuous part. The proof for the discontinuous part is completely analogous.

Step 1: Decomposition of the loading estimator:

First we start with the decomposition in Lemma D.1 that we get from substituting the definition of X into $\frac{1}{N}\hat{X}^{C\top}\hat{X}^C\hat{\Lambda}^C {V_{MN}^C}^{-1} = \hat{\Lambda}^C$. We choose H^C to set $\frac{1}{N}\Lambda^C {F^C}^\top F^C {\Lambda^C}^\top \hat{\Lambda}^C = \Lambda^C H V_{MN}^C$.

$$\begin{aligned} V_{MN}^C\left(\hat{\Lambda}_i^C - {H^C}^\top \Lambda_i^C\right) =& \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\hat{\Lambda}_k^C e_{j,k} e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq\alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\hat{\Lambda}_k^C {\Lambda_k^C}^\top F_j^C e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq\alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\hat{\Lambda}_k^C e_{j,k} {F_j^C}^\top \mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda_i^C + R^C \end{aligned}$$

with

$$
\begin{aligned}
R^C = &+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \Lambda_k^D e_{j,k} F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^D e_{j,i} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^D F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^D \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^D \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^C F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{C^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&- \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{C^\top} \Lambda_i^C \\
=& o_p(1)
\end{aligned}
$$

The convergence rate of R^C would be straightforward if the truncations were in terms of F and e_i instead of X. However using the same argument as in Lemma H.1, we can conclude that under Assumption 3 at most for a finite number of indices i it holds that $F_j \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} - F_j \mathbb{1}_{\{\|F_j\|\leq \alpha\Delta_M^{\bar{\omega}}\}} = O_p\left(\frac{1}{\sqrt{\delta}}\right)$ for M sufficiently large and otherwise the difference is equal to 0. Likewise if there is no jump in F $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} = e_{j,i}\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ except for a finite number of indices. Hence, we have a similar decomposition for $\left(\hat{\Lambda}_i^C - H^{C^\top}\Lambda_i^C\right)$ as in Lemma D.1 using only truncated observations.

Step 2: $\hat{\Lambda}_i^C - H^{C^\top}\Lambda_i^C = O_p\left(\frac{1}{\sqrt{\delta}}\right)$:

We need to show Lemmas D.2 and D.3 for the truncated observations. Note that Proposition C.1 does not hold any more because the truncated residuals are not necessarily local martingales any more. For this reason we obtain a lower convergence rate of $O_p\left(\frac{1}{\sqrt{\delta}}\right)$ instead of $O_p\left(\frac{1}{\delta}\right)$. The statement follows from a repeated use of Lemma H.1.

Step 3: Convergence of $\hat{F}_T^C - H^{C^{-1}} F_T^C$:

We try to extend Theorem 4 to the truncated variables. By abuse of notation I denote by $\Lambda^\top F \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ the matrix with elements $\Lambda_i^\top F_j \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ and similarly $e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ is the matrix with elements $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$.

$$\begin{aligned}
\hat{F}^C - F^C {H^{C}}^{-1\top} =& \frac{1}{N}\hat{X}^C\hat{\Lambda}^C - F^C {H^{C}}^{-1\top} \\
=& \frac{1}{N}\left(F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{\Lambda^C}^\top + F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{\Lambda^D}^\top + e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\hat{\Lambda}^C - F^C {H^{C}}^{-1\top} \\
=& \frac{1}{N}F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{\Lambda^C}^\top\hat{\Lambda}^C - F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{H^{C}}^{-1\top} + F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{H^{C}}^{-1\top} \\
&+ \frac{1}{N}F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{\Lambda^D}^\top\hat{\Lambda}^C + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\hat{\Lambda}^C - F^C {H^{C}}^{-1\top} \\
=& \frac{1}{N}F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left({\Lambda^C}^\top - {H^{C}}^{-1\top}\hat{\Lambda}^{C\top}\right)\hat{\Lambda}^C + \left(F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} - F^C\right){H^{C}}^{-1\top} \\
&+ F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\frac{1}{N}{\Lambda^D}^\top\Lambda^C H^C\right) + F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{1}{N}{\Lambda^D}^\top\left(\hat{\Lambda}^C - \Lambda^C H^C\right) \\
&+ \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda}^C - \Lambda^C H^C\right) + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C H^C.
\end{aligned}$$

Using the result $\hat{\Lambda}_i^C - {H^C}^\top\Lambda_i^C = O_p\left(\frac{1}{\sqrt{\delta}}\right)$ and a similar reasoning as in Lemma H.1, we conclude that

$$\hat{F}_T^C - {H^C}^{-1}F_T^C = o_p(1) + \left(\frac{1}{N}{\Lambda^D}^\top\Lambda^C H^C\right)^\top F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} + \frac{1}{N}{H^C}^\top{\Lambda^C}^\top e_T^\top\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$$

The term $F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\frac{1}{N}{\Lambda^D}^\top\Lambda^C H^C\right)$ goes to zero only if F^D has no drift term or Λ^D is orthogonal to Λ^C. Note that in general F^D can be written as a pure jump martingale and a finite variation part. Even when F^D does not jump its value does not equal zero because of the finite variation part. Hence in the limit $F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ estimates the drift term of F^D. A similar argument applies to $\frac{1}{N}e_T\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C H^C$. By definition e_i are local martingales. If the residuals also have a jump component, then this component can be written as a pure jump process minus its compensator, which is a predictable finite variation process. The truncation estimates the continuous part of e_i which is the continuous martingale plus the compensator process of the jump martingale. Hence, in the limit $e_i\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ is not martingale any more. In particular the weighted average of the compensator drift process does not vanish. In conclusion, if the jump factor process has a predictable finite variation part or more than finitely many residual terms have a jump component, there will be a predictable finite variation process as bias for the continuous factor estimator.

Step 4: Convergence of quadratic covariation:

The quadratic covariation estimator of the estimator $\hat{F}^C$ with another arbitrary process Y is

$$\sum_{j=1}^{M} \hat{F^C}_j Y_j = H^{C^{-1}} \sum_{j=1}^{M} F_j^C Y_j + o_p(1) + \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}$$
$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}.$$

The first term converges to the desired quantity. Hence, we need to show that the other two terms go to zero.

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} =& \frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T \\
&+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \left(\mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - \mathbb{1}_{\{|e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} \right) \\
&+ \frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} \left(\sum_{j=1}^{M} e_{j,i} Y_j \mathbb{1}_{\{|e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - [e_i^C, Y]_T \right)
\end{aligned}$$

The last two term are $o_p(1)$ by a similar argument as in Lemma H.1. Applying the Cauchy Schwartz inequality and Assumption 1 to the first term yields

$$\|\frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T\|^2 \leq \|\frac{1}{N^2} H^{C^\top} \Lambda^{C^\top} [e^C, e^C]_T \Lambda^C H^C\| \cdot \|[Y, Y]_T\| = O_p\left(\frac{1}{N}\right)$$

Thus Assumption 1 implies that $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T = O_p\left(\frac{1}{\sqrt{N}}\right)$. The last result follows from that fact that the quadratic covariation of a predictable finite variation process with a semimartingale is zero and $F_j^D \mathbb{1}_{\{\|F_j^D\| \leq \alpha \Delta_M^{\bar{\omega}}\}}$ converges to a predictable finite variation term:

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} &= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{\|F_j^D\| \leq \alpha \Delta_M^{\bar{\omega}}\}} + o_p(1) \\
&= o_p(1)
\end{aligned}$$

□

I Estimation of the Number of Factors

Lemma I.1. Weyl's eigenvalue inequality

For any $M \times N$ matrices Q_i we have

$$\lambda_{i_1+...+i_K-(K-1)}\left(\sum_{k=1}^{K} Q_k\right) \leq \lambda_{i_1}(Q_1) + ... + \lambda_{i_K}(Q_K)$$

where $1 \leq i_1, ..., i_K \leq \min(N, M))$, $1 \leq i_1 + ... + i_K - (K-1) \leq \min(N, M)$ and $\lambda_i(Q)$ denotes the ith largest singular value of the matrix Q.

Proof. See Theorem 3.3.16 in Horn and Johnson (1991). □

Lemma I.2. Bound on non-systematic eigenvalues
Assume Assumption 1 holds and $O\left(\frac{N}{M}\right) \leq O(1)$. Then

$$\lambda_k(X^\top X) \leq O_p(1) \qquad \textit{for } k \geq K+1.$$

Proof. Note that the singular values of a symmetric matrix are equal to the eigenvalues of this matrix. By Weyl's inequality for singular values in Lemma I.1 we obtain

$$\lambda_k(X) \leq \lambda_k(F\Lambda^\top) + \lambda_1(e).$$

As $\lambda_k(F\Lambda^\top) = 0$ for $k \geq K+1$, we conclude

$$\lambda_k(X^\top X) \leq \lambda_1(e^\top e) \qquad \text{for } k \geq K+1$$

Now we need to show that $\lambda_k(e^\top e) \leq O_p(1)$ $\forall k \in [1, N]$. We start with a decomposition

$$\begin{aligned}\lambda_k(e^\top e) &= \lambda_k(e^\top e - [e,e] + [e,e]) \\ &\leq \lambda_1(e^\top e - [e,e]) + \lambda_k([e,e]).\end{aligned}$$

By Assumption 1 $[e, e]$ has bounded eigenvalues, which implies $\lambda_k([e, e]) = O_p(1)$.

Denote by V the eigenvector of the largest eigenvalue of $(e^\top e - [e,e])$.

$$\begin{aligned}
\lambda_1\left(e^\top e - [e,e]\right) &= V^\top \left(e^\top e - [e,e]\right) V \\
&= \sum_{i=1}^N \sum_{l=1}^N V_i \left(e_i^\top e_l - [e_i,e_l]\right) V_l \\
&\le \left(\sum_{i=1}^N \left(\sum_{l=1}^N (e_i^\top e_l - [e_i,e_l]) V_l\right)^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^N V_i^2\right)^{\frac{1}{2}} \\
&\le \left(\sum_{i=1}^N \left(\sum_{l=1}^N (e_i^\top e_l - [e_i,e_l]) V_l\right)^2\right)^{\frac{1}{2}}
\end{aligned}$$

as V is an orthonormal vector. Apply Proposition C.1 with $Y = e_i$ and $\bar{Z} = \sum_{l=1}^N e_l V_l$. Note that $[\bar{Z},\bar{Z}] = V^\top [e,e] V$ is bounded. Hence

$$\sum_{l=1}^N \left(e_i^\top e_l - [e_i,e_l]\right) V_l = O_p\left(\frac{1}{\sqrt{M}}\right).$$

Therefore

$$\lambda_1\left(e^\top e - [e,e]\right) = \left(\sum_{i=1}^N O_p\left(\frac{1}{M}\right)\right)^{\frac{1}{2}} \le O_p\left(\frac{\sqrt{N}}{\sqrt{M}}\right) \le O_p(1).$$

□

Lemma I.3. Bound on systematic eigenvalues
Assume Assumption 1 holds and $O\left(\frac{N}{M}\right) \le O(1)$. *Then*

$$\lambda_k(X^\top X) = O_p(N) \qquad \textit{for } k = 1, ..., K$$

Proof. By Weyl's inequality for singular values in Lemma I.1:

$$\lambda_k(F\Lambda^\top) \le \lambda_k(X) + \lambda_1(-e)$$

By Lemma I.2 the last term is $\lambda_1(-e) = -\lambda_N(e) = O_p(1)$. Therefore

$$\lambda_k(X) \ge \lambda_k(F\Lambda^\top) + O_p(1)$$

which implies $\lambda_k(X^\top X) \ge O_p(N)$ as $\left(F^\top F \frac{\Lambda^\top \Lambda}{N}\right)$ has bounded eigenvalues for $k = 1, ..., K$. On the other hand

$$\lambda_k(X) \leq \lambda_k(F\Lambda^\top) + \lambda_1(e)$$

and $\lambda_1(e) = O_p(1)$ implies $\lambda_k(X^\top X) \leq O_p(N)$ for $k = 1, ..., K$. $\square$

Lemma I.4. Bounds on truncated eigenvalues
Assume Assumptions 1 and 3 hold and $O\left(\frac{N}{M}\right) \leq O(1)$. Set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}_{j,i}^C = X_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}_{j,i}^D = X_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}\}}$. Then

$$\begin{aligned}
&\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right) = O_p(N) && k = 1, ..., K_C \\
&\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right) \leq O_p(1) && k = K_C + 1, ..., N \\
&\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right) = O_p(N) && k = 1, ..., K_D \\
&\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right) \leq O_p(1) && k = K_D + 1, ..., N
\end{aligned}$$

where K^C is the number of factors that contain a continuous part and K^D is the number of factors that have a jump component.

Proof. By abuse of notation the vector $\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e$ has the elements $\mathbb{1}_{\{|e_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}e_{j,i}$. e^C is the continuous martingale part of e and e^D denotes the jump martingale part.
Step 1: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)\right) \leq O_p(1)$ for $k = 1, ..., N$.
By Lemma I.1 it holds that

$$\lambda_k(\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e) \leq \lambda_1(\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e - e^C) + \lambda_k(e^C)$$

Lemma I.2 applied to e^C implies $\lambda_k(e^C) \leq O_p(1)$. The difference between the continuous martingale part of e and the truncation estimator $\mathbb{1}_{\{|e| \leq \alpha\Delta_M^{\bar{\omega}}\}}e - e^C$ equals a drift term from the jump martingale part plus a vector with finitely many elements that are of a small order:

$$\mathbb{1}_{\{|e_i| \leq \alpha\Delta_M^{\bar{\omega}}\}}e_i - e_i^C = b_{e_i} + d_{e_i}$$

where b_{e_i} is a vector that contains the finite variation part of the jump martingales which is classified as continuous and d_{e_i} is a vector that contains the negative continuous part $-e_{j,i}^C$ for the increments j that are correctly classified as jumps and hence are set to zero in $\mathbb{1}_{\{|e_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}e_{j,i}$. Using the results of Mancini (2009) we have $\mathbb{1}_{\{e_{j,i}^D = 0\}} = \mathbb{1}_{\{|e_{j,i}| \leq \alpha\Delta^{\bar{\omega}}\}}$ almost surely for sufficiently large M and hence we can identify all the increments that contain jumps. Note, that by

Assumption 3 we have only finitely many jumps for each time interval and therefore d_{e_i} has only finitely many elements not equal to zero. By Lemma I.1 we have

$$\lambda_1(\mathbb{1}_{\{|e|\leq\alpha\Delta_M^{\bar{\omega}}\}}e - e^C) \leq \lambda_1(b_e) + \lambda_1(d_e)$$

It is well-known that the spectral norm of a symmetric $N \times N$ matrix A is bounded by N times its largest element: $\|A\|_2 \leq N \max_{i,k} |A_{i,k}|$. Hence

$$\lambda_1(b_e^\top b_e) \leq N \cdot \max_{k,i} |b_{e_i}^\top b_{e_k}| \leq N \cdot O_p\left(\frac{1}{M}\right) \leq O_p\left(\frac{N}{M}\right) \leq O_p(1)$$

where we have use the fact that the increments of a finite variation term are of order $O_p\left(\frac{1}{M}\right)$. Similarly

$$\lambda_1\left(d_e^\top d_e\right) \leq N \cdot \max_{k,i} |d_{e_i}^\top d_{e_k}| \leq N \cdot O_p\left(\frac{1}{M}\right) \leq O_p\left(\frac{N}{M}\right) \leq O_p(1)$$

as d_{e_i} has only finitely many elements that are not zero and those are of order $O_p\left(\frac{1}{\sqrt{M}}\right)$.

Step 2: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}e\right)\right) \leq O_p(1)$ for $k = 1, ..., N$.
Here we need to show that the result of step 1 still holds, when we replace $\mathbb{1}_{\{|e_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ with $\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$. It is sufficient to show that

$$\lambda_1\left(e\mathbb{1}_{\{|e|\leq\alpha\Delta^{\bar{\omega}}\}} - e\mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\right) := \lambda_1(h) = O_p(1)$$

As by Assumption 3 only finitely many elements of h are non-zero and those are of order $O_p\left(\frac{1}{\sqrt{M}}\right)$, it follows that

$$\lambda_1(h) \leq N \max_{k,i} |h_i^\top h_k| \leq O_p\left(\frac{N}{M}\right) \leq O_p(1).$$

Step 3: To show: $\lambda_k(\hat{X}^{C\top}\hat{X}^C) \leq O_p(1)$ for $k \geq K_C + 1$.
By definition the estimated continuous movements are

$$\hat{X}^C = F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C + F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top} + e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$$

where $F^{\text{pure jump}}$ denotes the pure jump factors that do not have a continuous component and $\Lambda^{\text{pure jump}}$ are the corresponding loadings. By Weyl's inequality for singular values in Lemma I.1 we have

$$\lambda_1\left(\hat{X}^C\right) \leq \lambda_1\left(F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C\right) + \lambda_1\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right) + \lambda_1\left(e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)$$

For $k \geq K+1$ the first term vanishes $\lambda_1\left(F^C \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C\right)=0$ and by step 2 the last term is $\lambda_1\left(e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)=O_p(1)$. The second term can be bounded by

$$\lambda_1\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right)^2 \leq \|\Lambda^{\text{pure jump}\top}\Lambda^{\text{pure jump}}\|_2^2 \cdot \|\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)^\top F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\|_2^2$$

The first factor is $\|\Lambda^{\text{pure jump}\top}\Lambda^{\text{pure jump}}\|_2^2 = O(N)$, while the truncated quadratic covariation in the second factor only contains the drift terms of the factors denoted by b_{F^D} which are of order $O_p\left(\frac{1}{M}\right)$:

$$\left\|\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)^\top F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right\|_2^2 \leq \|b_{F^D}{}^\top b_{F^D}\|_2^2 \leq O_p\left(\frac{1}{M}\right)$$

Step 4: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}e\right)\right)\leq O_p(1)$ for $k=1,...,N$.
We decompose the truncated error terms into two components.

$$\lambda_k(\mathbb{1}_{\{|e|>\alpha\Delta_M^{\bar{\omega}}\}}e) > \lambda_1(\mathbb{1}_{\{|e|>\alpha\Delta_M^{\bar{\omega}}\}}e-e^D)+\lambda_k(e^D).$$

By Proposition C.1 the second term is $O_p(1)$. For the first term we can apply a similar logic as in step 1. Then we use the same arguments as in step 2.
Step 5: To show: $\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right)=O_p(N)$ for $k=1,...,K^C$.
By Lemma I.1 the first K^C singular values satisfy the inequality

$$\lambda_k\left(F^C\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{C\top}\right)\leq\lambda_k\left(\hat{X}^C\right)+\lambda_1\left(-F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right)+\lambda_1\left(-e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right).$$

Hence by the previous steps

$$\lambda_k\left(\hat{X}^C\right)\geq\lambda_k\left(F^C\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{C\top}\right)+O_p(1).$$

By Assumption 1 for $k=1,...,K_C$

$$\lambda_k^2\left(F^C\Lambda^{C\top}\right)=\lambda_k\left(F^{C\top}F^C\frac{\Lambda^{C\top}\Lambda^C}{N}\right)N=O_p(N).$$

On the other hand

$$\lambda_k\left(F^C\mathbb{1}_{\{|X|\leq \alpha\Delta^{\bar{\omega}}\}}\Lambda^{C\top}-F^C\Lambda^{C\top}\right)^2\leq O_p\left(\frac{N}{M}\right)\leq O_p(1)$$

where we have used the fact that the difference between a continuous factor and the truncation estimator applied to the continuous part is just a finite number of terms of order $O_p\left(\frac{1}{\sqrt{M}}\right)$. Hence

$$\lambda_k^2\left(F^C \mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\Lambda^{C\top}\right) = O_p(N)$$

Similarly we get the reverse inequality for $\hat{X}^C$:

$$\lambda_k\left(\hat{X}^C\right) \leq \lambda_k\left(F^C \mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\Lambda^{C\top}\right) + \lambda_1\left(F^{\text{pure jump}} \mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right) + \lambda_1\left(e\mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\right)$$

which yields

$$O_p(N) \leq \lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right) \leq O_p(N)$$

Step 6: To show: $\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right) = O_p(N)$ for $k = 1, ..., K^D$.
Analogous to step 5. □

Proof of Theorem 9:

Proof. I only prove the result for $\hat{K}(\gamma)$. The results for $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ follow exactly the same logic.
Step 1: ER_k for $k = K$
By Lemmas I.2 and I.3 the eigenvalue ratio statistic for $k = K$ is asymptotically

$$ER_k = \frac{\lambda_K + g}{\lambda_{K+1} + g} = \frac{\frac{O_p(N)}{g} + 1}{\frac{\lambda_{K+1}}{g} + 1} = \frac{\frac{O_p(N)}{g} + 1}{o_p(1) + 1} = O_p\left(\frac{N}{g}\right) \to \infty$$

Step 2: ER_k for $k \geq K + 1$

$$ER_k = \frac{\lambda_k + g}{\lambda_{k+1} + g} = \frac{\frac{\lambda_k}{g} + 1}{\frac{\lambda_{k+1}}{g} + 1} = \frac{o_p(1) + 1}{o_p(1) + 1} = 1 + o_p(1).$$

Step 3: To show: $\hat{K}(\gamma) \xrightarrow{p} K$
As ER_k goes in probability to 1 for $k \geq K + 1$ and grows without bounds for $k = K$, the probability for $ER_k > 1$ goes to zero for $k \geq K + 1$ and to 1 for $k = K$.
Remark: Although it is not needed for this proof, note that for $k = 1, ..., K - 1$

$$ER_k = \frac{\lambda_k + g}{\lambda_{k+1} + g} = \frac{O_p(N) + g}{O_p(N) + g} = \frac{O_p(1) + \frac{g}{N}}{O_p(1) + \frac{g}{N}} = O_p(1).$$

$\square$

Proof of Proposition 1:

Proof. Apply Theorem L.7 to $\frac{1}{\sqrt{M}}X_{j,i} = \frac{1}{\sqrt{M}}F_j\Lambda_i^\top + \frac{1}{\sqrt{M}}e_{j,i}$. Note that $\frac{1}{\sqrt{M}}e$ can be written as $\frac{1}{\sqrt{M}}e = A\epsilon$ with $\epsilon_{j,i}$ being *i.i.d.* $(0,1)$ random variables with finite fourth moments. $\square$

J Identifying the Factors

Proof of Theorem 11:

Proof. Define

$$B = \begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix} \qquad \hat{B} = \begin{pmatrix} \hat{F}^\top \hat{F} & \hat{F}^\top G \\ G^\top \hat{F} & G^\top G \end{pmatrix} \qquad B^* = \begin{pmatrix} H^{-1}F^\top F H^{-1^\top} & H^{-1}F^\top G \\ G^\top F H^{-1^\top} & G^\top G \end{pmatrix}.$$

As the trace is a linear function it follows that $\sqrt{M}\left(trace(B) - trace(\hat{B})\right) \xrightarrow{p} 0$ if $\sqrt{M}(B - \hat{B}) \xrightarrow{p} 0$. By assumption H is full rank and the trace of B is equal to the trace of B^*. Thus it is sufficient to show that $\sqrt{M}(\hat{B} - B^*) \xrightarrow{p} 0$. This follows from

(i) $\sqrt{M}\left((\hat{F}^\top \hat{F})^{-1} - (H^{-1}F^\top F H^{-1^\top})^{-1}\right) \xrightarrow{p} 0$

(ii) $\sqrt{M}\left(\hat{F}^\top G - H^{-1}F^\top G\right) \xrightarrow{p} 0$.

We start with (i). As

$$(\hat{F}^\top \hat{F})^{-1} - (H^{-1}F^\top F H^{-1^\top})^{-1} = (\hat{F}^\top \hat{F})^{-1}\left(H^{-1}F^\top F H^{-1^\top} - \hat{F}^\top \hat{F}\right)\left(H^{-1}F^\top F H^{-1^\top}\right)^{-1}$$

it is sufficient to show

$$\sqrt{M}\left(H^{-1}F^\top F H^{-1^\top} - \hat{F}^\top \hat{F}\right) = \sqrt{M}H^{-1}F^\top(FH^{-1^\top} - \hat{F}) + \sqrt{M}(H^{-1}F^\top - \hat{F}^\top)\hat{F} \xrightarrow{p} 0$$

It is shown in the proof of Theorem 4 that

$$\hat{F} - FH^{-1^\top} = \frac{1}{N}F(\Lambda - \hat{\Lambda}H^{-1})^\top \hat{\Lambda} + \frac{1}{N}e(\hat{\Lambda} - \Lambda H) + \frac{1}{N}e\Lambda H.$$

Hence the first term equals

$$-H^{-1}F^\top(\hat{F} - FH^{-1^\top}) = \frac{1}{N}H^{-1}F^\top F(\Lambda - \hat{\Lambda}H^{-1})^\top \hat{\Lambda} + \frac{1}{N}H^{-1}F^\top e(\hat{\Lambda} - \Lambda H) + \frac{1}{N}H^{-1}F^\top e\Lambda H$$

Lemmas D.2 and E.1 applied to the first summand yield $\frac{1}{N}H^{-1}F^\top F(\Lambda - \hat{\Lambda}H^{-1})^\top \hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$. Lemmas C.1 and D.2 provide the rate for the second summand as $\frac{1}{N}H^{-1}F^\top e(\hat{\Lambda} - \Lambda H) = O_p\left(\frac{1}{\delta}\right)$.

Lemma C.1 bounds the third summand: $\frac{1}{N} H^{-1} F^\top e \Lambda H = O_p\left(\frac{1}{\sqrt{NM}}\right)$.

For the second term note that

$$\left(H^{-1} F^\top - \hat{F}^\top\right) \hat{F} = \left(H^{-1} F^\top - \hat{F}^\top\right) \left(F H^{-1^\top} - \hat{F}\right) + \left(H^{-1} F^\top - \hat{F}^\top\right) F H^{-1^\top}$$

Based on Lemmas D.2 and E.1 it is easy to show that $\left(H^{-1} F^\top - \hat{F}^\top\right) \left(F H^{-1^\top} - \hat{F}\right) = O_p\left(\frac{1}{\delta}\right)$.

Term (ii) requires the additional assumptions on G:

$$\left(\hat{F}^\top - H^{-1} F^\top\right) G = \left(\frac{1}{N} \hat{\Lambda}^\top \left(\Lambda - \hat{\Lambda} H^{-1}\right) F^\top G + \frac{1}{N} \left(\hat{\Lambda} - \Lambda H\right)^\top e^\top G + \frac{1}{N} H^\top \Lambda^\top e^\top G.\right.$$

By Lemma E.1 it follows that $\left(\frac{1}{N} \hat{\Lambda}^\top \left(\Lambda - \hat{\Lambda} H^{-1}\right)\right) F^\top G = O_p\left(\frac{1}{\delta}\right)$. Now let's first assume that G is independent of e. Then Proposition C.1 applies and $\frac{1}{N} H^\top \Lambda e^\top G = O_p\left(\frac{1}{\sqrt{NM}}\right)$. Otherwise assume that $G = \frac{1}{N} \sum_{i=1}^N X_i w_i^\top = F \frac{1}{N} \sum_{i=1}^N \Lambda_i w_i^\top + \frac{1}{N} \sum_{i=1}^N e_i w_i^\top$. Proposition C.1 applies to

$$\frac{1}{N} H^\top \Lambda e^\top F \left(\frac{1}{N} \sum_{i=1}^N \Lambda_i w_i^\top\right) = O_p\left(\frac{1}{\sqrt{NM}}\right)$$

and

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{N} H^\top \Lambda^\top \left(e^\top e_i - [e, e_i]\right)\right) w_i^\top = O_p\left(\frac{1}{\sqrt{NM}}\right)$$

separately. As by Assumption 2

$$\sum_{i=1}^N \frac{1}{N^2} H^\top \Lambda^\top [e, e_i] w_i^\top = \frac{1}{N^2} \left(\sum_{i=1}^N \sum_{k=1}^N H^\top \Lambda_k [e_k, e_i] w_i^\top\right) = O_p\left(\frac{1}{N}\right)$$

the statement in (ii) follows. The distribution result is a consequence of the delta method for the function

$$f\left(\begin{pmatrix} [F, F] \\ [F, G] \\ [G, F] \\ [G, G] \end{pmatrix}\right) = trace\left([F, F]^{-1} [F, G] [G, G]^{-1} [G, F]\right)$$

which has the partial derivates

$$\begin{aligned}\frac{\partial f}{\partial [F,F]} &= -\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F][F,F]^{-1}\right)^\top \\ \frac{\partial f}{\partial [F,G]} &= [F,F]^{-1}[F,G][G,G]^{-1} \\ \frac{\partial f}{\partial [G,F]} &= [G,G]^{-1}[G,F][F,F]^{-1} \\ \frac{\partial f}{\partial [G,G]} &= -\left([G,G]^{-1}[G,F][F,F]^{-1}[F,G][G,G]^{-1}\right)^\top\end{aligned}$$

Hence

$$\sqrt{M}\left(\hat{\bar{\rho}} - \bar{\rho}\right) = \xi^\top \sqrt{M}\left(vec\left(\begin{pmatrix}[F,F] & [F,G] \\ [G,F] & [G,G]\end{pmatrix} - B\right)\right) + \sqrt{M}\cdot trace\left(B^* - \hat{B}\right)$$

The last term is $O_p\left(\frac{\sqrt{M}}{\delta}\right)$ which goes to zero by assumption. □

Proof of Theorem 12:

Proof. The theorem is a consequence of Theorem 11 and Section 6.1.3 in Aït-Sahalia and Jacod (2014). □

K Microstructure Noise

Lemma K.1. Limits of extreme eigenvalues
Let Z be a $M \times N$ double array of independent and identically distributed random variables with zero mean and unit variance. Let $S = \frac{1}{M}Z^\top Z$. Then if $\mathbb{E}[|Z_{11}|^4] < \infty$, as $M \to \infty$, $N \to \infty$, $\frac{N}{M} \to c \in (0,1)$, we have

$$\begin{aligned}\lim \lambda_{\min}(S) &= \left(1-\sqrt{c}\right)^2 \qquad a.s. \\ \lim \lambda_{\max}(S) &= \left(1+\sqrt{c}\right)^2 \qquad a.s.\end{aligned}$$

where $\lambda_i(S)$ denotes the ith eigenvalue of S.

Proof. See Bai and Yin (1993) □

Proof of Theorem 10:

Proof. **Step 1:** To show: $\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \lambda_1\left(\frac{\epsilon^\top \epsilon}{N}\right) + \lambda_1\left(\frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}\right)$
This is an immediate consequence of Weyl's eigenvalue inequality Lemma I.1 applied to the

matrix

$$\frac{(e+\epsilon)^\top(e+\epsilon)}{N} = \frac{e^\top e}{N} + \frac{\epsilon^\top \epsilon}{N} + \frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}.$$

Step 2: To show: $\lambda_1\left(\frac{e^\top\epsilon}{N}+\frac{\epsilon^\top e}{N}\right) = O_p\left(\frac{1}{N}\right)$
Let V be the eigenvector for the largest eigenvalue of $\frac{e^\top\epsilon}{N}+\frac{\epsilon^\top e}{N}$. Then

$$\begin{aligned}\lambda_1\left(\frac{e^\top\epsilon}{N}+\frac{\epsilon^\top e}{N}\right) &= V^\top\frac{e^\top\epsilon}{N}V + V^\top\frac{\epsilon^\top e}{N}V\\ &= 2\frac{1}{N}\sum_{j=1}^M\sum_{i=1}^N\sum_{k=1}^N V_i\epsilon_{j,i}e_{j,i}V_k.\end{aligned}$$

Define $\bar\epsilon_j = \sum_{i=1}^N V_i\epsilon_{j,i}$ and $\bar e_j = \sum_{k=1}^N V_k e_{j,k}$. As can be easily checked $\bar\epsilon_j\bar e_j$ form a martingale difference sequence and hence we can apply Burkholder's inequality in Lemma L.2:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^M\bar\epsilon_j\bar e_j\right)^2\right] &\le C\sum_{j=1}^M\mathbb{E}\left[\bar\epsilon_j^2\bar e_j^2\right] \le C\sum_{j=1}^M\mathbb{E}\left[\bar\epsilon_j^2\right]\mathbb{E}\left[\bar e_j^2\right] \le \frac{C}{M}\sum_{j=1}^M\mathbb{E}\left[\bar\epsilon_j^2\right]\\ &\le \frac{C}{M}\mathbb{E}\left[\left(\sum_{i=1}^N V_i\epsilon_{j,i}\right)^2\right] \le \frac{C}{M}\sum_{i=1}^N V_i^2\mathbb{E}\left[\epsilon_{j,i}^2\right] \le C.\end{aligned}$$

We have used the Burkholder inequality to conclude $\mathbb{E}\left[\bar e_j^2\right] \le CV^\top\mathbb{E}[\Delta_j\langle e,e\rangle]V \le \frac{C}{M}$. This shows that $V^\top\frac{e^\top\epsilon}{N}V = O_p\left(\frac{1}{N}\right)$.

Step 3: To show: $\lambda_1\left(\frac{\epsilon^\top\epsilon}{N}\right) \le \frac{1}{c}\left(1+\sqrt{c}\right)^2\lambda_1(B^\top B)\sigma_\epsilon^2 + o_p(1)$
Here we define B as

$$B = \begin{pmatrix}1 & -1 & 0 & \cdots & 0\\ 0 & 1 & -1 & \cdots & 0\\ \vdots & & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & 1 & -1\\ 0 & 0 & \cdots & 0 & 1\end{pmatrix}$$

and note that $\epsilon = B\tilde\epsilon$ (up to the boundaries which do not matter asymptotically). Now we can split the spectrum into two components:

$$\lambda_1\left(\frac{\epsilon^\top\epsilon}{N}\right) = \lambda_1\left(\frac{\tilde\epsilon^\top B^\top B\tilde\epsilon}{N}\right) \le \lambda_1\left(\frac{\tilde\epsilon^\top\tilde\epsilon}{N}\right)\lambda_1\left(B^\top B\right).$$

By Lemma K.1 it follows that

$$\lambda_1\left(\frac{\tilde{\epsilon}^\top\tilde{\epsilon}}{N}\right) = \frac{1}{c}\left((1+\sqrt{c})^2\sigma_\epsilon^2\right) + o_p(1).$$

Step 4: To show: $\sigma_\epsilon^2 \leq \frac{c}{\left(1-\sqrt{c}\right)^2}\frac{\lambda_s\left(\frac{Y^\top Y}{N}\right)}{\lambda_{s+K}(B^\top B)} + o_p(1)$
Weyl's inequality for singular values Lemma I.1 implies

$$\lambda_{s+K}(e+\epsilon) \leq \lambda_{K+1}(F\Lambda^\top) + \lambda_s(Y) \leq \lambda_s(Y)$$

as $\lambda_{K+1}(F\Lambda^\top) = 0$. Lemma A.6 in Ahn and Horenstein (2013) says that if A and B are $N \times N$ positive semidefinite matrices, then $\lambda_i(A) \leq \lambda_i(A+B)$ for $i = 1, ..., N$. Combining this lemma with step 2 of this proof, we get

$$\lambda_{s+K}\left(\frac{\epsilon^\top\epsilon}{N}\right) \leq \lambda_s\left(\frac{Y^\top Y}{N}\right)$$

Lemma A.4 in Ahn and Horenstein (2013) yields

$$\lambda_N(\tilde{\epsilon}^\top\tilde{\epsilon})\lambda_{s+K}(B^\top B) \leq \lambda_{s+K}(\epsilon^\top\epsilon)$$

Combining this with lemma K.1 gives us

$$\frac{1}{c}\left((1-\sqrt{c})^2\sigma_\epsilon^2\right)\lambda_{s+K}(B^\top B) + o_p(1) \leq \lambda_s\left(\frac{Y^\top Y}{N}\right)$$

Solving for σ_ϵ^2 yields the statement.

Step 5: To show: $\lambda_s(B^\top B) = 2\left(1+\cos\left(\frac{s+1}{N+1}\pi\right)\right)$
$B^\top B$ is a symmetric tridiagonal Toeplitz matrix with 2 on the diagonal and -1 on the off-diagonal. Its eigenvalues are well-known and equal $2 - 2\cos\left(\frac{N-s}{N+1}\pi\right) = 2\left(1+\cos\left(\frac{s+1}{N+1}\pi\right)\right)$.

Step 6: Combining the previous steps.

$$\begin{aligned}\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) &\leq \left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2\frac{2\left(1+\cos\left(\frac{2}{N+1}\pi\right)\right)}{2\left(1+\cos\left(\frac{s+1+K}{N}\pi\right)\right)}\lambda_s\left(\frac{Y^\top Y}{N}\right) + o_p(1)\\ &\leq \left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2\frac{2}{1+cos\left(\frac{s+K+1}{N}\pi\right)}\lambda_s\left(\frac{Y^\top Y}{N}\right) + o_p(1)\end{aligned}$$

for all $s \in [K+1, N_K]$. Here we have used the continuity of the cosinus function. □

L Collection of Limit Theorems

Theorem L.1. Localization procedure
Assume X is a d-dimensional Itô semimartingale on $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ defined as

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \mathbb{1}_{\{\|\delta\|\leq 1\}} \delta(s,x)(\mu - \nu)(ds, dx) + \int_0^t \int_E \mathbb{1}_{\{\|\delta\|>1\}} \delta(s,x)\mu(ds, dx)$$

where W is a d-dimensional Brownian motion and μ is a Poisson random measure on $\mathbb{R}_+ \times E$ with $(E, \mathbb{E})$ an auxiliary measurable space on the space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ and the predictable compensator (or intensity measure) of μ is $\nu(ds, dx) = ds \times v(dx)$.

The volatility σ_t is also a d-dimensional Itô semimartingale of the form

$$\begin{aligned}\sigma_t =& \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s + \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|\leq 1\}} \tilde{\delta}(s,x)(\mu - \nu)(ds, dx) \\ &+ \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}} \tilde{\delta}(s,x)\mu(ds, dx)\end{aligned}$$

where W' is another Wiener process independent of (W, μ). Denote the predictable quadratic covariation process of the martingale part by $\int_0^t a_s ds$ and the compensator of $\int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}} \tilde{\delta}(s,x)\mu(ds, dx)$ by $\int_0^t \tilde{a}_s ds$.

Assume local boundedness denoted by Assumption H holds for X:

1. *The process b is locally bounded and cádlág.*
2. *The process σ is càdlàg.*
3. *There is a localizing sequence τ_n of stopping times and, for each n, a deterministic nonnegative function Γ_n on E satisfying $\int \Gamma_n(z)^2 v(dz) < \infty$ and such that $\|\delta(\omega, t, z)\| \wedge 1 \leq \Gamma_n(z)$ for all (ω, t, z) with $t \leq \tau_n(\omega)$.*

The volatility process also satisfies a local boundedness condition denoted by Assumption K:

1. *The processes $\tilde{b}$, a and $\tilde{a}$ are locally bounded and progressively measurable*
2. *The process $\tilde{\sigma}$ is càdlàg or càglàd and adapted*

We introduce a global boundedness condition for X denoted by Assumption SH: Assumption H holds and there are a constant C and a nonnegative function Γ on E such that

$$\begin{aligned}&\|b_t(\omega)\| \leq C \qquad \|\sigma_t(\omega)\| \leq C \qquad \|X_t(\omega)\| \leq C \qquad \|\delta(\omega, t, z)\| \leq \Gamma(z) \\ &\Gamma(z) \leq C \qquad \int \Gamma(z)^2 v(dz) \leq C.\end{aligned}$$

Similarly a global boundedness condition on σ is imposed and denoted by Assumption SK: We have Assumption K and there are a constant and a nonnegative function Γ on E, such that

Assumption SH holds and also

$$\|\tilde{b}_t(\omega)\| \leq C \qquad \|\tilde{\sigma}_t(\omega)\| \leq C \qquad \|a_t(\omega)\| \leq C \qquad \|\tilde{a}_t(\omega)\| \leq C.$$

The processes $U^n(X)$ and $U(X)$ are subject to the following conditions, where X and X' are any two semimartingales that satisfy the same assumptions and S is any $(\mathfrak{F}_t)$-stopping time: $X_t = X'_t$ a.s. $\forall t < S \Rightarrow$

- $t < S \Rightarrow U^n(X)_t = U^n(X')_t$ *a.s.*
- *the $\mathfrak{F}$-conditional laws of $(U(X)_t)_{t<S}$ and $(U(X')_t)_{t<S}$ are a.s. equal.*

The properties of interest for us are either one of the following properties:

- *The processes $U^n(X)$ converge in probability to $U(X)$*
- *The variables $U^n(X)_t$ converge in probability to $U(X)_t$*
- *The processes $U^n(X)$ converge stably in law to $U(X)$*
- *The variables $U^n(X)_t$ converge stably in law to $U(X)_t$.*

If the properties of interest hold for Assumption SH, then they also hold for Assumption H. Likewise, if the properties of interest hold for Assumption SK, they also hold for Assumption K.

Proof. See Lemma 4.4.9 in Jacod and Protter (2012). □

Theorem L.2. Central limit theorem for quadratic variation
Let X be an Itô semimartingale satisfying Definition 1. Then the $d \times d$-dimensional processes $\bar{Z}^n$ defined as

$$\bar{Z}^n_t = \frac{1}{\sqrt{\Delta}} \left([X,X]^n_t - [X,X]_{\Delta[t/\Delta]} \right)$$

converges stably in law to a process $\bar{Z} = \left(\bar{Z}^{ij}\right)_{1 \leq i,j \leq d}$ defined on a very good filtered extension $(\tilde{\Omega}, \tilde{\mathfrak{F}}, (\tilde{\mathfrak{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ of $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ and which, conditionally on $\mathfrak{F}$, is centered with independent increments and finite second moments given by

$$\begin{aligned}\mathbb{E}\left[\bar{Z}^{ij}_t \bar{Z}^{kl}_t | \mathfrak{F}\right] = &\frac{1}{2} \sum_{s \leq t} \Bigg(\Delta X^i_s \Delta X^k_s (c^{jl}_{s-} + c^{jl}_s) + \Delta X^i_s \Delta X^l_s (c^{jk}_{s-} + c^{jk}_s) \\ &+ \Delta X^j_s \Delta X^k_s (c^{il}_{s-} + c^{il}_s) + \Delta X^j_s \Delta X^l_s (c^{ik}_{s-} + c^{ik}_s) \Bigg) + \int_0^t \left(c^{ik}_s c^{jl}_s + c^{il}_s c^{jk}_s \right) ds\end{aligned}$$

with $c_t = \sigma_t^\top \sigma_t$. This process $\bar{Z}$ is $\mathfrak{F}$-conditionally Gaussian, if the process X and σ have no common jumps.

Moreover, the same is true of the process $\frac{1}{\sqrt{\Delta}}([X,X]^n - [X,X])$*, when* X *is continuous, and otherwise for each* t *we have the following stable convergence of variables*

$$\frac{1}{\sqrt{\Delta}}([X,X]^n_t - [X,X]_t) \overset{L-s}{\to} \bar{Z}_t.$$

Proof. See Jacod and Protter (2013) Theorem 5.4.2. □

Theorem L.3. Consistent Estimation of Covariance in Theorem L.2
We want to estimate

$$D_t = \sum_{s \leq t} |\Delta X|^2 (\sigma_{s-} + \sigma_s)$$

Let X *be an Itô semimartingale satisfying Definition 1. In addition for some* $0 \leq r < 1$ *it satisfies the stronger assumption that there is a localizing sequence* τ_n *of stopping times and for each* n *a deterministic nonnegative function* Γ_n *on* E *satisfying* $\int \Gamma_n(z)\lambda(dz) < \infty$ *and such that* $\|\delta(\omega,t,z)\|^r \wedge 1 \leq \Gamma_n(z)$ *for all* (ω,t,z) *with* $t \leq \tau_n(\omega)$*.*

Assume that $\frac{1}{2(2-r)} \leq \bar{\omega} < \frac{1}{2}$ *and let* u_M *be proportional to* $\frac{1}{M^{\bar{\omega}}}$*. Choose a sequence* k_n *of integers with the following property:*

$$k \to \infty, \qquad \frac{k}{M} \to 0$$

We set

$$\hat{\sigma}(\bar{\omega})_j = \frac{M}{k} \sum_{m=0}^{k-1} (\Delta_{j+m} X)^2 \mathbb{1}_{\{|\Delta_{j+m} X| \leq u_M\}}$$

Define $\hat{D} = \sum_{j=k+1}^{[t \cdot M]-k} |\Delta_j X|^2 \mathbb{1}_{\{|\Delta_j X| > u_M\}} \cdot (\hat{\sigma}_{j-k} + \hat{\sigma}_{j+1})$ *Then*

$$\hat{D} \overset{p}{\to} D$$

Proof. See Theorem A.7 in Aït-Sahalia and Jacod (2014). □

Lemma L.1. Martingale central limit theorem
Assume $Z_n(t)$ *is a sequence of local square integrable martingales and* Z *is a Gaussian martingale with quadratic characteristic* $\langle Z, Z \rangle$*. Assume that for any* $t \in (t, T]$

1. $\int_0^t \int_{|z|>\epsilon} z^2 \nu^n(ds, dx) \overset{p}{\to} 0 \ \forall \epsilon \in (0,1]$
2. $[Z_n, Z_n]_t \overset{p}{\to} [Z, Z]_t$

Then $Z_n \overset{D}{\to} Z$ *for* $t \in (0, T]$*.*

Proof. See Lipster and Shiryayev (1980) □

Theorem L.4. Martingale central limit theorem with stable convergence
Assume $X^n = \{(X_t^n, \mathfrak{F}_t^n; 0 \le t \le 1\}$ are cádlág semimartingales with $X_0^n = 0$ and histories $\mathfrak{F}^n = \{\mathfrak{F}_t^n; 0 \le t \le 1\}$.

$$X_t^n = X_0^n + \int_0^t b_s^{X^n} ds + \int_0^t \sigma_s^{X^n} dW_s + \int_0^t \int_E \mathbb{1}_{\{\|x\|\le 1\}}(\mu^{X^n} - \nu^{X^n})(ds,dx) + \int_0^t \int_E \mathbb{1}_{\{\|x\|>1\}}\mu^{X^n}(ds,dx)$$

We require the nesting condition of the $\mathfrak{F}^n$: There exists a sequence $t_n \downarrow 0$ such that

1. $\mathfrak{F}_{t_n}^n \subseteq \mathfrak{F}_{t_{n+1}}^{n+1}$
2. $\bigvee_n \mathfrak{F}_{t_n}^n = \bigvee_n \mathfrak{F}_1^n$

Define C = { g: continuous real functions, zero in a neighborhood of zero, with limits at ∞ }
Suppose

1. *D is dense in $[0,1]$ and $1 \in D$.*
2. *X is a quasi-left continuous semimartingale.*
3. (a) $\forall t \in D\ \sup_{s\le t} |b_s^{X^n} - b_s^X| \xrightarrow{p} 0.$
   (b) $\forall t \in D\ \langle X^{nc}\rangle_t + \int_0^t \int_{|x|<1} x^2 d\nu^{X^n} - \sum_{s\le t} |\Delta b_s^{X^n}|^2 \xrightarrow{p} \langle X^c\rangle_t + \int_0^t \int_{|x|<1} x^2 \nu^X(ds,dx).$
   (c) $\forall t \in D\ \forall g \in C\ \int_0^t \int_{\mathbb{R}} g(x)\nu^{X^n}(ds,dx) \xrightarrow{p} \int_0^t \int_{\mathbb{R}} g(x)\nu^X(ds,dx).$

Then

$$X^n \overset{L-s}{\to} X$$

in the sense of stable weak convergence in the Skorohod topology.

Proof. See Theorem 1 in Feigin (1984). □

Lemma L.2. Burkholder's inequality for discrete martingales
Consider a discrete time martingale $\{S_j, \mathfrak{F}_j, 1 \le j \le M\}$. Define $X_1 = S_1$ and $X_j = S_j - S_{j-1}$ for $2 \le j \le M$. Then, for $1 < p < \infty$, there exist constants C_1 and C_2 depending only on p such that

$$C_1 \mathbb{E}\left[\sum_{j=1}^M X_i^2\right]^{p/2} \le \mathbb{E}|S_M|^p \le C_2 \mathbb{E}\left[\sum_{j=1}^M X_j^2\right]^{p/2}.$$

Proof. See Theorem 2.10 in Hall and Heyde (1980). □

Lemma L.3. Burkholder-Davis-Gundy inequality

For each real $p \geq 1$ there is a constant C such that for any local martingale M starting at $M_0 = 0$ and any two stopping times $S \leq T$, we have

$$E\left[\sup_{t\in\mathbb{R}^+:S\leq t\leq T} |M_t - M_S|^p|\mathfrak{F}_S\right] \leq CE\left[([M,M]_T - [M,M]_S)^{p/2}\,|\mathfrak{F}_S\right].$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma L.4. Hölder's inequality applied to drift term

Consider the finite variation part of the Itô semimartingale defined in Definition 1. We have

$$\sup_{0\leq u\leq s} \|\int_T^{T+u} b_r dr\|^2 \leq s\int_T^{T+s} \|b_u\|^2 du.$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma L.5. Burkholder-Davis-Gundy inequality for continuous martingales

Consider the continuous martingale part of the Itô semimartingale defined in Definition 1. There exists a constant C such that

$$E\left[\sup_{0\leq u\leq s} \|\int_T^{T+u} \sigma_r dW_r\|^2|\mathfrak{F}_T\right] \leq CE\left[\int_T^{T+s} \|\sigma_u\|^2 du|\mathfrak{F}_T\right]$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma L.6. Burkholder-Davis-Gundy inequality for purely discontinuous martingales

Suppose that $\int_0^t \int \|\delta(s,z)\|^2 v(dz)ds < \infty$ for all t, i.e. the process $Y = \delta \star (\mu - \nu)$ is a locally square integrable martingale. There exists a constant C such that for all finite stopping times T and $s > 0$ we have

$$E\left[\sup_{0\leq u\leq s} \|Y_{T+u} - Y_T\|^2|F_T\right] \leq CE\left[\int_T^{T+s}\int \|\delta(u,z)\|^2 v(dz)du|\mathfrak{F}_T\right].$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Theorem L.5. Detecting Jumps

Assume X is an Itô-semimartingale as in Definition 1 and in addition has only finite jump activity, i.e. on each finite time interval there are almost surely only finitely many bounded *jumps. Denote* $\Delta_M = \frac{T}{M}$ *and take a sequence* v_M *such that*

$$v_M = \alpha \Delta_M^{\bar{\omega}} \qquad \text{for some } \bar{\omega} \in \left(0, \frac{1}{2}\right) \text{ and a constant } \alpha > 0.$$

Our estimator classifies an increment as containing a jump if

$$\Delta_j X > v_M.$$

Denote by $I_M(1) < ... < I_M(\hat{R})$ *the indices* j *in* $1, ..., M$ *such that* $\Delta_j X > v_M$*. Set* $\hat{T}_{jump}(q) = I_M(q) \cdot \Delta_M$ *for* $q = 1, ..., \hat{R}$*. Let* $R = \sup(q : T_{jump}(q) \leq T)$ *be the number of jumps of* X *within* $[0, T]$*. Then we have*

$$\mathbb{P}\left(\hat{R} = R, T_{jump}(q) \in (\hat{T}_{jump}(q) - \Delta_M, \hat{T}_{jump}(q)] \quad \forall q \in \{1, ..., R\}\right) \to 1$$

Proof. See Theorem 10.26 in Aït-Sahalia and Jacod (2014). □

Theorem L.6. Estimation of continuous and discontinuous quadratic covariation
Assume X *is an Itô-semimartingale as in Definition 1 and in addition has only finite jump activity, i.e. on each finite time interval there are almost surely only finitely many bounded jumps. Denote* $\Delta_M = \frac{T}{M}$ *and take some* $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ *and a constant* $\alpha > 0$*. Define the continuous component of* X *by* X^C *and the discontinuous part by* X^D*. Then*

$$\sum_{j=1}^{M} X_j^2 \mathbb{1}_{\{|X_j| \leq \alpha \Delta_M^{\bar{\omega}}\}} = [X^C, X^C] + O_p\left(\frac{1}{\sqrt{M}}\right)$$
$$\sum_{j=1}^{M} X_j^2 \mathbb{1}_{\{|X_j| > \alpha \Delta_M^{\bar{\omega}}\}} = [X^D, X^D] + O_p\left(\frac{1}{\sqrt{M}}\right).$$

Proof. See Theorem A.16 in Aït-Sahalia and Jacod (2014). Actually they make a much stronger statement and characterize the limiting distribution of the truncation estimators. □

Theorem L.7. Onatski estimator for the number of factors
Assume a factor model holds with

$$X = F \Lambda^\top + e$$

where X *is a* $M \times N$ *matrix of* N *cross-sectional units observed over* M *time periods.* Λ *is a* $N \times K$ *matrix of loadings and the factor matrix* F *is a* $M \times K$ *matrix. The idiosyncratic* *component e is a $M \times N$ matrix and can be decomposed as*

$$e = A\epsilon B$$

with a $M \times M$ matrix A, a $N \times N$ matrix B and a $M \times N$ matrix ϵ.

Define the eigenvalue distribution function of a symmetric $N \times N$ matrix S as

$$\mathcal{F}^S(x) = 1 - \frac{1}{N}\#\{i \leq N : \lambda_i(S) > x\}$$

where $\lambda_1(S) \geq ... \geq \lambda_N(S)$ are the ordered eigenvalues of S. For a generic probability distribution having bounded support and cdf $\mathcal{F}(x)$, let $u(\mathcal{F})$ be the upper bound of the support, i.e. $u(\mathcal{F}) = \min\{x : (x) = 1\}$. The following assumptions hold:

1. *For any constant $C > 0$ and $\delta > 0$ there exist positive integers N_0 and M_0 such that for any $N > N_0$ and $M > M_0$ the probability that the smallest eigenvalue of $\frac{\Lambda^\top \Lambda}{N}\frac{F^\top F}{M}$ is below C is smaller than δ.*
2. *For any positive integers N and M, the decomposition $e = A\epsilon B$ holds where*
   (a) *$\epsilon_{t,i}$, $1 \leq i \leq N$, $1 \leq t \leq M$ are i.i.d. and satisfy moment conditions $\mathbb{E}[\epsilon_{t,i}] = 0$, $\mathbb{E}[\epsilon_{t,i}^2] = 1$ and $\mathbb{E}[\epsilon_{t,i}^4] < \infty$.*
   (b) *$\mathcal{F}^{AA^\top}$ and $\mathcal{F}^{BB^\top}$ weakly converge to probability distribution functions $\mathcal{F}_A$ and $\mathcal{F}_B$ respectively as N and M go to infinity.*
   (c) *Distributions $\mathcal{F}_A$ and $\mathcal{F}_B$ have bounded support, $u(\mathcal{F}^{AA^\top}) \to u(\mathcal{F}_A) > 0$ and $u(\mathcal{F}^{BB^\top}) \to u(\mathcal{F}_B) > 0$ almost surely as N and M go to infinity. $\liminf_{\delta \to 0} \delta^{-1} \int_{u(\mathcal{F}_A)-\delta}^{u(\mathcal{F}_A)} d\mathcal{F}_A(\lambda) = k_A > 0$ and $\liminf_{\delta \to 0} \delta^{-1} \int_{u(\mathcal{F}_B)-\delta}^{u(\mathcal{F}_B)} d\mathcal{F}_B(\lambda) = k_B > 0$.*
3. *Let $M(N)$ be a sequence of positive integers such that $\frac{N}{M(N)} \to c > 0$ as $N \to \infty$.*
4. *Let ϵ either have Gaussian entries or either A or B are a diagonal matrix*

Then as $N \to \infty$, we have

1. *For any sequence of positive integers $r(N)$ such that $\frac{r(N)}{N} \to 0$ as $N \to \infty$ and $r(N) > K$ for large enough N the $r(N)$th eigennvalue of $\frac{X^\top X}{NM}$ converges almost surely to $u(\mathcal{F}^{c,A,B})$ where $\mathcal{F}^{c,A,B}$ is the distribution function defined in Onatski (2010).*
2. *The K-th eigenvalue of $\frac{X^\top X}{NM}$ tends to infinity in probability.*
3. *Let $\{K_{max}^N, N \in \mathbb{N}\}$ be a slowly increasing sequence of real numbers such that $K_{max}^N/N \to 0$ as $N \to \infty$. Define*

   $$\hat{K}^\delta = \max\{i \leq K_{max}^N : \lambda_i - \lambda_{i+1} \geq \delta\}$$

   For any fixed $\delta > 0$ $\hat{K}(\delta) \to K$ in probability as $N \to \infty$.

Proof. See Onatski (2010). $\square$

References

Ahn, S. C., and A. R. Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203–1227.

Aït-Sahalia, P. A. Mykland, Y., and L. Zhang, 2005a, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* 18, 351–416.

Aït-Sahalia, P. A. Mykland, Y., and L. Zhang, 2005b, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394–1411.

Aït-Sahalia, Y., 2004, Disentangling diffusion from jumps, *Journal of Financial Economics* 74, 487–528.

Aït-Sahalia, Y., J. Fan, and D. Xiu, 2010, High-frequency estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* 105, 1504–1516.

Aït-Sahalia, Y., and J. Jacod, 2009a, Estimating the degree of activity of jumps in high frequency data, *Annals of Statistics* 37, 2202–2244.

Aït-Sahalia, Y., and J. Jacod, 2009b, Testing for jumps in a discretely observed process, *Annals of Statistics* 37, 184–222.

Aït-Sahalia, Y., and J. Jacod, 2014, *High-Frequency Financial Econometrics* (New Jersey: Princeton University Press).

Aït-Sahalia, Y., and D. Xiu, 2015a, Principal component analysis of high frequency data, *Working paper* .

Aït-Sahalia, Y., and D. Xiu, 2015b, Principal component estimation of a large covariance matrix with high-frequency data, *Working paper* .

Aldous, D. G., and G. K. Eagleson, 1978, On mixing and stability of limit theorems, *Annal of Probability* 6, 325–331.

Alessi, L., M. Barigozzi, and M. Capassoc, 2010, Improved penalization for determining the number of factors in approximate factor models, *Statistics and Probability Letters* 80, 1806–1813.

Amengual, D., and M.W. Watson, 2007, Consistent estimation of the number of dynamic factors in a large n and t panel, *Journal of Business and Economic Statistics* 25, 91–96.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, An empirical investigation of continuous-time equity return models, *Journal of Finance* 57, 1239–1284.

Andersen, T.G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 42, 42–55.

Anderson, T. W., 1963, Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics* 34, 122–148.

Anderson, T.W., 1984, *An Introduction to Multivariate Statistical Analysis* (Wiley).

Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135–171.

Bai, J., and S. Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Bai, J., and S. Ng, 2006, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* 507–537.

Bai, Jushan, and Serena Ng, 2008, *Large dimensional factor analysis* (Now Publishers Inc).

Bai, Z. D., and Y. Q. Yin, 1993, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, *The Annals of Probability* 21, 1275–1294.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2011, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* 162, 149–169.

Barndorff-Nielsen, O.E., and N. Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society* 253–280.

Barndorff-Nielsen, O.E., and N. Shephard, 2004a, Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics, *Econometrica* 72, 885–925.

Barndorff-Nielsen, O.E., and N. Shephard, 2004b, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2, 1–48.

Barndorff-Nielsen, O.E., and N. Shephard, 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* 4, 1–30.

Barndorff-Nielsen, O.E., N. Shephard, and M Winkel, 2006, Limit theorems for multipower variation in the presence of jumps, *Stochastic Processes and their Applications* 116, 796–806.

Bibinger, M., and L. Winkelmann, 2014, Econometrics of co-jumps in high-frequency data with noise, *Journal of Econometrics* 184, 361–378.

Bollerslev, T., T.H. Law, and G. Tauchen, 2008, Risk, jumps, and diversification, *Journal of Financial Econometrics* 144, 234–256.

Bollerslev, T., S. Z. Li, and V. Todorov, 2013, Jump tails, extreme dependencies and the distribution of stock returns., *Journal of Econometrics* 172, 307–324.

Bollerslev, T., S. Z. Li, and V. Todorov, 2015a, Roughing up beta: Continuous vs. discontinuous betas, and the cross section of expected stock returns, *Working paper* .

Bollerslev, T., and V. Todorov, 2010, Jumps and betas: A new theoretical framework for disentangling and estimating systematic risks, *Journal of Econometrics* 157, 220–235.

Bollerslev, T., V. Todorov, and L. Xu, 2015b, Tail risk premia and return predictability., *Journal of Financial Economics* Forthcoming.

Chamberlain, G., 1988, Asset pricing in multiperiod securities markets, *Econometrica* 56, 1283–1300.

Chamberlain, G., and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Connor, G., and R. Korajczyk, 1988, Risk and return in an equilibrium APT: Application to a new test methodology, *Journal of Financial Economics* 21, 255–289.

Connor, G., and R. Korajczyk, 1993, A test for the number of factors in an approximate factor model,, *Journal of Finance* 58, 1263–1291.

Duffie, D., J. Pan, and K. J. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fan, J., A. Furger, and D. Xiu, 2014, Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data, *Working paper* .

Fan, J., Y. Li, and K. Yu, 2012, Vast volatility matrix estimation using high-frequency data for portfolio selection, *Journal of the American Statistical Association* 107, 412–428.

Fan, L., Y. Liao, and M. Mincheva, 2013, Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society* 75, 603–680.

Feigin, P. D., 1985, Stable convergence of semimartingales, *Stochastic Processes and their Applications* 19, 125–134.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2000, The generalized dynamic-factor model: Identification and estimation, *REVIEW* 82, 540–554.

Hall, P., and C.C. Heyde, 1980, *Martingale Limit Theory and its Application* (Academic Press).

Hallin, M., and R. Liska, 2007, The generalized dynamic factor model: Determining the number of factors, *Journal of the American Statistical Association* 102, 603–617.

Hansen, P., and A. Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–161.

Harding, M., 2013, Estimating the number of factors in large dimensional factor models, *Journal of Econometrics* .

Hayashi, T., and N. Yoshida, 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* 11, 359–379.

Horn, R. A., and C. R. Johnson, 1991, *Topics in Matrix Analysis* (Cambridge: Cambridge University Press).

Jacod, J., 2008, Asymptotic properties of realized power variations and related functionals of semimartingales, *Stochastic Processes and their Applications* 118, 517–559.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter, 2009, Microstructure noise in the continuous case: The pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249–2276.

Jacod, J., and P. Protter, 2012, *Discretization of Processes* (Heidelberg: Springer).

Jacod, J., and A. N. Shiryaev, 2002, *Limit Theorems for Stochastic Processes* (Springer).

Kapetanios, G., 2010, A testing procedure for determining the number of factors in approximate factor models, *Journal of Business and Economic Statistics* 28, 397–409.

Lee, S. S., and P. A. Mykland, 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* 21, 2535–2563.

Lunde, A., N. Shephard, and K. Sheppard, 2014, Econometric analysis of vast covariance matrices using composite realized kernels, *Working paper* .

Magnus, J. R., 1988, *Linear Structures* (Oxford University Press).

Mancini, C, 2009, Non parametric threshold estimation for models with stochastic diffusion coefficient and jumps, *Scandinavian Journal of Statistics* 42–52.

Marcenko, V. A., and L. A. Pastur, 1967, Distribution of eigenvalues for some sets of random matrices,, *Math. USSR-Sbornik* 457–483.

Onatski, A., 2010, Determining the number of factors from empirical distribution of eigenvalues, *Review of Economic and Statistics* 92, 1004–1016.

Paul, D., and J. Silverstein, 2008, No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix, *Working paper* .

Pelger, M., 2015, Understanding systematic risk: A high-frequency approach, *Working paper* .

Podolskij, M., and M. Vetter, 2009, Bipower-type estimation in a noisy diffusion setting, *Stochastic Processes and their Applications* 11, 2803–2831.

Prigent, J., 2003, *Weak Convergence of Financial Markets* (Heidelberg: Springer).

Protter, P.E, 2004, *Stochastic Integration and Differential Equations* (Springer-Verlag).

Rogers, C., and D. Williams, 2000, *Diffusions, Markov Processes and Martingales*, volume 2, second edition (Cambridge: Cambridge University Press).

Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Silverstein, J., and S. Choi, 1995, Analysis of the limiting spectral distribution of large dimensional random matrices, *Journal of Multivariate Analysis* 54, 295–309.

Stock, J., and M. Watson, 2002a, Macroeconomic forecasting using diffusion indexes,, *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J. H., and M. W. Watson, 2002b, Forecasting using principal components from a large number of predictors, *Journal of American Statistical Association* 97, 1167–1179.

Tao, M., Y. Wang, and X. Chen, 2013a, Fast convergence rates in estimating large volatility matrices using high-frequency financial data, *Econometric Theory* 29, 838–856.

Tao, M., Y. Wang, and H. H. Zhou, 2013b, Optimal sparse volatility matrix estimation for high dimensional Itô processes with measurement errors, *Annals of Statistics* 41, 1816–1864.

Todorov, V., 2009, Estimation of continuous-time stochastic volatility models with jumps using high-frequency data, *Journal of Econometrics* 148, 131–148.

Wang, Y., and J. Zhou, 2010, Vast volatility matrix estimation for high-frequency financial data, *Annals of Statistics* 38, 943–978.

Zhang, L., 2011, Estimating covariation: Epps effect, microstructure noise, *Journal of Econometrics* 160, 33–47.