"SOUND AND FURY": NONLINEAR FUNCTIONALS OF VOLATILITY MATRIX IN THE PRESENCE OF JUMP AND NOISE

BY RICHARD Y. CHEN[1]

[1] *richard.chen.ryc@gmail.com*

This paper resolves a pivotal open problem on nonparametric inference for nonlinear functionals of volatility matrix. Multiple prominent statistical tasks can be formulated as functionals of volatility matrix, yet a unified statistical theory of general nonlinear functionals based on noisy data remains challenging and elusive. Nonetheless, this paper shows it can be achieved by combining the strengths of pre-averaging, jump truncation and nonlinearity bias correction. In light of general nonlinearity, bias correction beyond linear approximation becomes necessary. Resultant estimators are nonparametric and robust over a wide spectrum of stochastic models. Moreover, the estimators can be rate-optimal and stable central limit theorems are obtained. The proposed framework lends itself conveniently to uncertainty quantification and permits fully feasible inference. With strong theoretical guarantees, this paper provides an inferential foundation for a wealth of statistical methods for noisy high-frequency data, such as realized principal component analysis, continuous-time linear regression, realized Laplace transform, generalized method of integrated moments and specification tests, hence extends current application scopes to noisy data which is more prevalent in practice.

1. Introduction. Functions of covariance matrix are the formulations in various eminent statistical applications. The inverse function leads to precision matrix, while a more prosaic example is the identity function (covariance matrix). Their applications include, for example, graphical models and discriminant analysis. Model selection and testing are oftentimes accomplished by functions of covariance matrix, as statistical models imply linear and nonlinear relationships among its elements and blocks. As for factor models and principal component analysis, eigenvalues or associated spectral functions are differentiable nonlinear functions of the corresponding covariance matrix, so are eigenvectors.

Accordingly, functionals of stochastic volatility matrix are critical in applications using time-dependent high-frequency data. High-frequency data often arises from financial transactions, neurological measurements, satellite weather recordings, gravitational wave detectors etc. To demonstrate a concrete application, we will study principal component analysis from the perspective of nonlinear functionals in Section 5; and we will present an empirical analysis of transaction data from TAQ (Trade and Quote) database in Section 6.

Extant literature on functionals of volatility matrix demands the assumption that there is no noise in the data. This no-noise assumption implies, as the sampling frequency becomes higher, consistent estimators converge to parameters of interest. However, this assumption

Keywords and phrases: nonlinear functionals, nonlinearity bias, stochastic volatility matrix, microstructure noise, jumps, stable central limit theorems, principal component analysis.

** This manuscript supersedes an earlier draft named "*Inference for Volatility Functionals of Multivariate Itô Semimartingales Observed with Jump and Noise*". The overlap with the previous draft consists of stochastic model in Section 2, rate-optimal estimator in Section 3, Theorem 1 in Section 4. What're new in this manuscript include statistical applications of the proposed methodology, a new family of estimators with a new set of hyper-parameters that facilitate applications, additional theorem and propositions that show consistency, asymptotic Gaussianity, finite-sample positive definiteness. Section 6 is new and we demonstrate a concrete financial application with large high-frequency dataset and discuss its empirical implication on portfolio construction.

imposes a restricting limitation on the type of data that we can use, as the empirical fact of high-frequency data tells us an opposing story of divergence at higher frequencies. Figure 1 presents how estimators behave as sampling interval Δ_n shrinks. The subscript n in Δ_n denotes sample size. Larger sample sizes come with smaller sampling intervals, i.e., $\Delta_n \to 0$ as $n \to \infty$. Strikingly the estimator for noiseless data does not converge for large sample sizes.

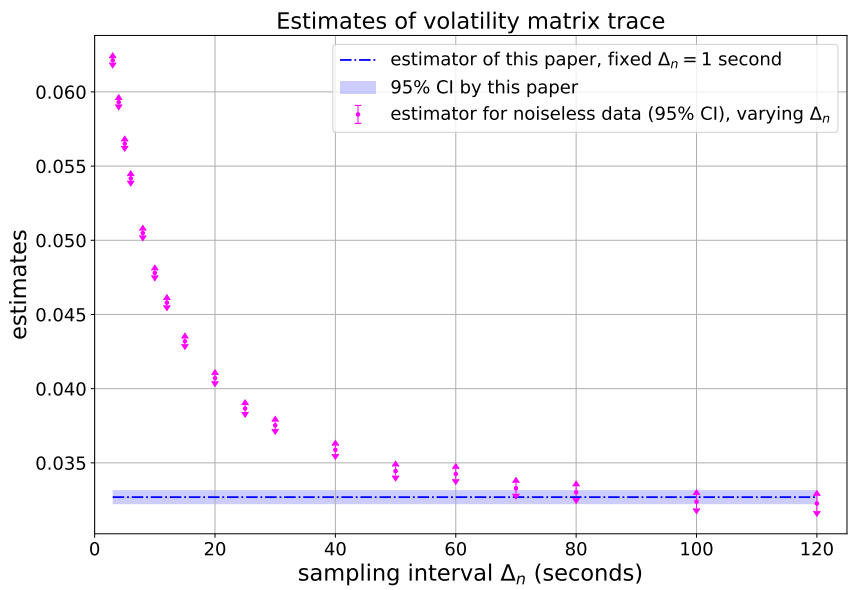


FIG 1. *Estimators applied to S&P 100 transaction data during 12/16/2019 - 12/20/2019*

Noise affects statistical significance and severely biases estimation. The presence of microstructure noise is commonly acknowledged. Solving the noise problem is vital to statistics of high-frequency data. The literature has called for a solution to this challenging open problem for years, e.g. [1, 2, 3, 4, 5, 6].[1] In this paper we provide a viable approach to this much sought-after solution.

Formally, functionals of volatility matrix are defined as

$$S(g)_t = \int_0^t g(c_s)\,\mathrm{d}s, \tag{1.1}$$

here t is finite; $g : \mathbb{R}^{d\times d} \mapsto \mathbb{R}^r$ is a twice continuously differentiable function on some convex cone of positive-definite matrices; c_s is stochastic volatility matrix, viz., instantaneous covariation of the continuous part of Itô semimartingale (e.g. jump-diffusion model). (1.1) is the inferential objective and it is of paramount statistical importance. By suitable choices of g, the quantity $S(g)_t$ can aid researchers and practitioners in uncovering valuable yet hidden information about cross sections and temporal evolutions from datasets of interest.

Indeed, a large amount of research efforts have been devoted to this kind of questions in the last twenty years. Early literature largely concentrated in stochastic volatility matrix and linear functionals; recent research advanced on general nonlinear functionals. It may come as

[1] For reference, see p. 1108 in [1]; p. 1617 in [2]; p. 38 in [3]; Section 6 in [4]; p. 159 in [5]; Remark 1 in [6].

no surprise that general functionals of stochastic volatility matrix provide a powerful unifying framework.

For linear functionals of stochastic volatility, the problem of microstructure noise has been resolved. Some influential approaches are two-scales realized volatility [7, 8], realized kernel method [9, 10], pre-averaging method [11, 12, 13, 14], quasi-maximum likelihood method [15, 16, 17], spectral local method of moments [18, 19].

For general nonlinear functionals, existing literature has absolved the inferential problem in the absence of microstructure noise. The pioneering work [20] put forward a statistical framework based on contiguity and local likelihood for nonlinear transformations of volatility. The seminal paper [21] proposed a plug-in inferential framework for nonlinear functionals, laid out the statistical theory, and spurred a long line of follow-up work on statistical methodologies and empirical applications. A more recent work [22] proposed a spectral method to circumvent asynchronous observations across dimensions when conducting statistical inference for nonlinear functionals of volatility matrices.

Detailed study of nonlinear functionals of statistical importance and resultant empirical results using noise-free data blossomed in recent years, some examples are

- realized Laplace transform [1, 23, 24];
- generalized method of integrated moments [2];
- continuous-time linear regressions [3];
- realized principal component analysis [4, 27];
- estimation of asymptotic variances, e.g. the quantity called quarticity [21];
- model specification tests [25, 26].

These aforementioned applications utilized high-frequency transaction data to answer statistical questions in economics and finance, such as model calibration [1, 23, 24], testing economic models [2, 25, 26], measuring financial risks [3, 4, 27], and uncertainty quantification [21]. Behind this powerful formulation and the wide range of applicability, however, there lurk difficult challenges.

A substantial challenge is nonlinearity. General nonlinearity induces high-order derivatives which contributes to the estimation error. Local linear approximation of nonlinear functionals only guarantees consistency and a suboptimal rate of convergence. In order to improve estimation accuracy and establish central limit theorems, bias correction of the high-order nonlinearity terms is indispensable, cf. [21, 28]. On one hand, direction correction of these nonlinearity biases requires computing gradients and Hessian matrices. On the other hand, nonlinearity bias correction requires careful tuning in nonparametric estimation. Creative and groundbreaking advancements have been made, under the no-noise assumption, to relax tuning restriction, facilitate bias correction and variance estimation by multiscale jackknife [5] and matrix calculus [26]. In situations where noise must be considered, notwithstanding one more layer of complexity, we find novel nonlinearity bias terms and a new tuning range that allows successful bias correction.

The second significant challenge is jumps. Discontinuity is frequently encountered in functional data. It is also a stylized fact in financial time series. In consequence, nonparametric jump-diffusion models are widely accepted by the literature. When estimating functionals of volatility matrix, the technique to handle discontinuity is jump truncation, which has been used in [5, 26, 21, 28]. However, jump and noise exhibit subtle interplay and delicate intricacy as documented in [29, 30]. Jump truncation for the purpose of volatility matrix functionals in the noisy setting is far from being straightforward. Despite this complexity, we establish a theoretical base that takes account of noise and jump simultaneously. It is one of the contributions of this paper.

Nevertheless, the central challenge is to find an effective statistical method and theory robust to microstructure noise. Figure 1 has demonstrated the curse of noise, especially in the

regime where the sampling interval Δ_n is small. As made manifest in Figure 2, Δ_n being a few seconds or smaller is typical for high-frequency financial data. Yet, because of the no-noise assumption in the extant literature on nonlinear functionals, the current available statistical methods require researchers to subsample high-frequency data so that Δ_n is between 1 minute to 5 minutes, e.g. [4, 31, 32]. The goal of subsampling is to mitigate the degree of noise accumulation, so that estimators designed for noise-free data are able to yield sound results, cf. Figure 1.

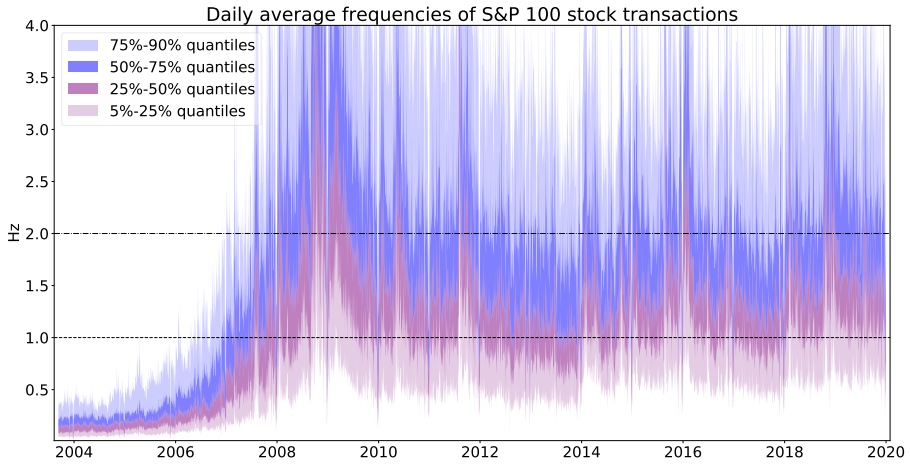


FIG 2. *Daily average frequencies of S&P 100 stock transactions*

In contrast to mitigation, we adopt a "bottom-up" approach, i.e., to start with a more realistic stochastic model that incorporates microstructure noise, then design noise-robust statistical methods, and then provide relevant statistical theory. We do not dispute the validity of subsampling as a mitigation, but we would like to emphasize that utilizing data of higher frequencies in a noise-robust fashion indeed gains statistical benefits, namely better estimation accuracy and sharper confidence intervals. It is an exemplar that data gives feedback to the model, and the attendant change delivers a better statistical performance.

The limitation imposed by the no-noise assumption on the amount of usable data is a pressing issue, and will be more so as we are observing explosions of data in the data-revolution age. The TAQ database of financial time series is no exception. The size of TAQ transaction data is depicted in Figure 3. TAQ is timestamped at millisecond latency and is subject to noise contamination. If equipped with a noise-robust statistical methodology, we can improve Δ_n from 5 minutes to, for example, 1 second. In this case there would immediately be 300 times more data at our disposal. Consequently this paper enables applications of volatility functionals to harvest data at much higher frequencies than the state-of-the-art methods allow.

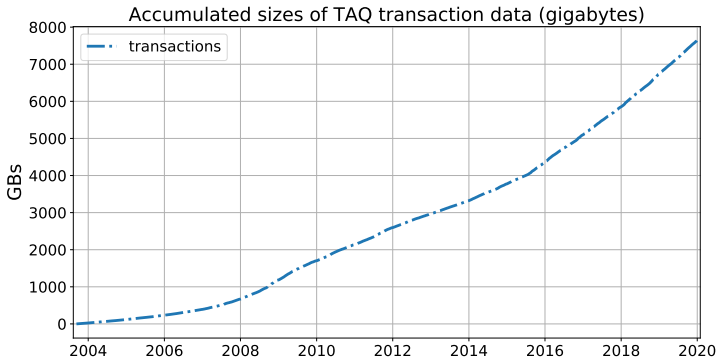


FIG 3. *Accumulated sizes of TAQ transaction data*

To cope with noise, we conduct local time-domain smoothing by the pre-averaging method [12, 13, 14]. To estimate general nonlinear functionals, we embed the pre-averaging method into the plug-in framework of [21]. In this direction, the technical advancement of this work is two-fold. One is to generalize the pre-averaging method from volatility matrix to its general functionals. The other is to extend the framework of functional inference to accommodate noisy data.

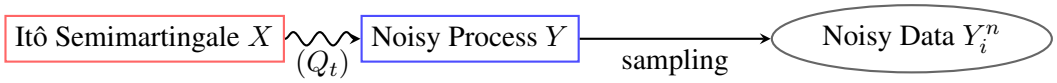
Besides nonlinearity bias, jumps, and noise discussed earlier, we need to tackle the following technicalities on the journey to asymptotic theory valuable to statistics.

- *Stochastic volatility*. Our model is nonparametric and can reproduce the salient characteristics of empirical financial data. A primary feature is stochastic volatility which models the volatility matrix as a stochastic process. However, to develop a robust and model-free methodology with satisfactory statistical guarantees, it becomes crucial to balance statistical error (dictated by the extent of local smoothing) and target error (attributable to time-varying parameters).
- *Dependence*. The pre-averaging method uses overlapping windows, hence the local moving averages are highly correlated for which standard central limit theorem is not valid. We use the "big block - small block" technique of [12]. It is instrumental to derive central limit theorems (hereafter CLT) for highly correlated high-frequency data, but in proof it induces technicality and involved notation.
- *Stable convergence*.[2] For statistical purpose, we need to establish stable convergence which is stronger than weak convergence (convergence in distribution). The reason is that sampling distributions of the estimators depend on the sample path of stochastic volatility matrix. We need a mode of convergence in which the estimators converge jointly with other variables, so that we can consistently estimate the asymptotic variances for the purpose of statistical inference.
- *Unbounded derivatives*. Some statistical applications, such as precision matrix estimation, specification tests, and linear regression, correspond to functionals with singularities in derivatives near the origin. The original framework [21] is not applicable in this situation. [3] proposed a spatial localization procedure that accommodates this type of functionals under the no-noise assumption. In the interest of applications using noisy data, we extend the spatial localization by establishing uniform convergence of the pre-averaging method plus jump truncation in the presence of discontinuity.

[2]Readers can read section 2.2.1 and section 2.2.2 in [33] for an exposition of stable convergence.

We state the stochastic model and our assumptions in Section 2. After presenting the appropriate range of tuning parameters, we show consistency, convergence rates, and associated CLT. There are two types of general results. The first satisfies the rate-optimal CLT, hence it is asymptotically superior; the second ensures positive semi-definiteness in finite samples, and in a sense requires less computation for statistical inference. We will give details about them in Section 3 and Section 4.
We focus on specific functionals which are the basis of principal component analysis (hereafter PCA) using noisy high-frequency data in Section 5. It has generated a lot of statistical interest and has been studies in, for example, [4, 27]. This paper significantly improves the accuracy of PCA either by being able to harness more data at higher frequencies, or by increasing the convergence rate. We also present recent examples among other interesting applications. Based on PCA of more than 16 years of noisy high-frequency data, a large-scale empirical analysis is presented in Section 6. Finally we conclude in Section 7. Proofs and simulations are in the appendices.

2. Setting. We describe our data generating mechanism and set up notation in this section. The model of high-frequency data with microstructure noise consists of three components: signal process, noise process, sampling scheme.



2.1. *Signal.* The underlying signal process is modeled by an Itô semimartingale X, i.e., X is a solution to the stochastic differential equation

$$X_t = X_0 + \int_0^t b_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s + J_t \tag{2.1}$$

where $b_s \in \mathbb{R}^d$, $\sigma_s \in \mathbb{R}^{d \times d'}$ with $d \leq d'$, b and c are adapted and càdlàg[3]; W is a d'-dimensional standard Brownian motion; J is a purely discontinuous process.
The mathematical description of J is relegated to (A.1), which can accommodate compound Poisson process, Lévy process etc. The stochastic volatility matrix is defined as $c_s = \sigma_s \sigma_s^{\mathrm{T}}$, which is stochastic process specified in (A.2).
The full signal model assumption is given in Appendix A, below is a summary of its features:

- the drift b has a smooth trajectory in the mean-square sense;
- the volatility c is a spatially constrained Itô semimartingale[4] such that c is locally bounded and well-conditioned;
- the celebrated leverage effect (correlation between price and volatility) is allowed, and is defined through the tensor product of σ and $\sigma^{(c)}$;
- J may exhibit finite and infinite activities, but has finite variation (i.e., finite-length trajectory) on compact sets;

These assumptions are necessary for the existence of CLT. The assumption that c is locally well-conditioned is for the applicability of some functionals of statistical interest, such as some specification tests and linear regression.
This signal model has no parametric restriction and allows time-varying parameters, e.g. the drift b_s and diffusion coefficient σ_s could be stochastic processes. Moreover it permits unrestricted non-stationarity and dependence among the state variables.

[3]"*continue à droite, limite à gauche*". It means "right continuous with left limits", i.e., the sample paths lie in the Skorokhod space.

[4]However, it is important to accommodate long-memory volatility model. The volatility functional inference in long-memory and noisy setting is an open question under investigation.

2.2. *Noise.* We adopt the probabilistic model in [12, 13] to describe noise. The advantage of this formalism is that it incorporates additive white noise, rounding error, the combination thereof as special cases. The gist is to suppose for $t > 0$ there is a probability transition kernel Q_t linking the value of a noisy process Y with that of X at time t. The abstract formalism is given in Appendix A.
The noise assumption is that under the conditional probability measure Q_t, the process $Y - X$ has zero mean, finite eighth moment, no serial correlation (though cross-sectional correlation is allowed). An example of this noisy process is given below.

EXAMPLE 1.

$$Y_t = f(X_t, e_t) \tag{2.2}$$

where $f : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d$ and e is a $\mathbb{R}^d$-valued perturbation process such that $\varepsilon_t = f(X_t, e_t) - X_t$ is centered white noise process conditional on X.

2.3. *Sampling.* This work treats regularly sampled observations and considers in-fill asymptotics.[5] Specifically, the samples are observed every Δ_n time units on a finite time interval $[0, t]$ where $n = \lfloor t/\Delta_n \rfloor$ is the sample size up to time t. As $\Delta_n \to 0$, $n \to \infty$ while t is fixed. We establish functional convergence and our statistical theory holds for any $t > 0$.
Throughout this paper, U_i^n is written for $U_{i\Delta_n}$ where U can be a process or filtration. For example, c_i^n denotes the stochastic volatility matrix at time $i\Delta_n$; let be $(\mathcal{F}_t)$ a filtration, $\mathcal{F}_i^n$ is the information up to time $i\Delta_n$. For any process U, $\Delta_i^n U$ represents the increment $U_i^n - U_{i-1}^n$.

2.4. *Notations.* For $r \in \mathbb{N}^+$, $\mathcal{C}^r(\mathcal{S})$ denotes the space of r-time continuously differentiable functions on the domain $\mathcal{S}$; $\mathcal{S}_d^+$ is the convex cone of $d \times d$ positive semidefinite matrices; $\mathbf{I}_r$ is the $r \times r$ identity matrix; $\|\cdot\|$ denotes a norm on vectors, matrices or tensors; given $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the largest integer no more than a; $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$; $a_n = O_p(b_n)$ means $\forall \epsilon > 0$, $\exists M > 0$ such that $\sup_n \mathbb{P}(a_n/b_n > M) < \epsilon$; $\mathbf{A}^{\mathrm{T}}$ is the transpose of the vector or matrix $\mathbf{A}$; for a multidimensional array, the entry index is written in the superscript, e.g., $X_t = (X_t^1, \cdots, X_t^d)^{\mathrm{T}}$, $\mathbf{A}^{jk}$ denotes the (j, k)-th entry in the matrix $\mathbf{A}$; $\otimes$ is the outer product, e.g., for two vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathrm{T}}$; $\partial_{jk} g$ and $\partial^2_{jk,lm} g$ denote the gradient and Hessian of g with respect to the (j, k)-th and (l, m)-th entries in its matrix argument; for a scalar function f define on a subset of $\mathbb{R}$, f' denotes its gradient. $\xrightarrow{\mathcal{L}-s(f)}$ (resp. $\xrightarrow{\mathcal{L}-s}$) denotes stable convergence in law of processes (resp. variables); $\xrightarrow{\mathcal{L}}$ denotes convergence in law (weak convergence); $\xrightarrow{u.c.p.}$ denotes uniform convergence in probability on compact sets; $\mathcal{MN}(0, V)$ is a mixed Gaussian distribution with zero mean and conditional covariance matrix V.

3. Methodology. The estimation methodology is comprised of five steps:

i. local moving averages of noisy data by a smoothing kernel;
ii. jump truncation operated on local moving averages;
iii. spot volatility estimator $\widehat{c}_i^n$'s for estimating c_i^n's;
iv. Riemann sum of $g(\widehat{c}_i^n)$'s for approximating the integral functional $\int g(c_s)\,\mathrm{d}s$;
v. nonlinearity bias correction.

[5] aka fixed-domain asymptotics, high-frequency asymptotics, small-interval asymptotics

The idea of local moving average is due to the pre-averaging method [12, 13]; the truncation is modified from (16.4.4) in [33]; the Riemann sum and bias correction are inspired by [21]. Though nonlinear bias does not impact on statistical consistency, we stress that we must carefully take into account the nonlinear terms since they lead to bias in the second order. To improve convergence rate and to establish CLT, nonlinearity bias correction is crucial.
The specific recipe is given next.

3.1. *Estimator of stochastic volatility matrix.* For the local moving averages, we choose a smoothing kernel φ such that

$$\begin{aligned} &\operatorname{supp}(\varphi) \subset (0,1),\ \textstyle\int_0^1 \varphi^2(s)\,\mathrm{d}s > 0 \\ &\varphi \in \mathcal{C} \text{ is piecewise } \mathcal{C}^1;\ \varphi' \text{ is piecewise Lipschitz} \end{aligned} \tag{3.1}$$

We choose an integer l_n as the number of observations in each smoothing window, define $\varphi_h^n = \varphi(h/l_n)$ and $\psi_n = \sum_{h=1}^{l_n-1}(\varphi_h^n)^2$. We associate the following quantities with the noisy process Y,

$$\begin{aligned} \overline{Y}_i^n &= (\psi_n)^{-1/2} \textstyle\sum_{h=1}^{l_n-1} \varphi_h^n\, \Delta_{i+h-1}^n Y \\ \widehat{Y}_i^n &= (2\psi_n)^{-1} \textstyle\sum_{h=0}^{l_n-1} (\varphi_{h+1}^n - \varphi_h^n)^2\, \Delta_{i+h}^n Y \otimes \Delta_{i+h}^n Y \end{aligned} \tag{3.2}$$

- $\overline{Y}_i^n$ is a local moving average of the noisy data Y_i^n's, and it serves as a proxy for $\Delta_i^n X$;
- $\widehat{Y}_i^n$ is an offset to the remaining noise in the outer product $\overline{Y}_i^n \otimes \overline{Y}_i^n$.

If there is jump(s) in the local window $(i\Delta_n, (i+l_n-1)\Delta_n]$, $\overline{Y}_i^n$ deviates from the target. We deem the moving average to be unreliable and veto it if $\|\overline{Y}_i^n\| > \nu_n$. The threshold of jump truncation is $\nu_n = \alpha\,\Delta_n^\rho$. We let α mirror the dimensionality and volatility level, and set the value of ρ according to the scaling property of Brownian motion (the choice of ρ is stated in (3.9)). There are some variants to this jump truncation. An alternative is elementwise truncation $|\overline{Y}_i^{r,n}| > \alpha_i^r \Delta_n^\rho$ where α_i^r is related to the local volatility of the r-th component. We recommend jump truncation tuned to optimize finite-sample performance, and it is used in our simulation and data analysis. We should emphasize that these variants share the same asymptotic behavior and satisfy the same asymptotic theory.
Based on these ingredients, we choose k_n such that $k_n > l_n$, and design the following estimator of stochastic volatility matrix:

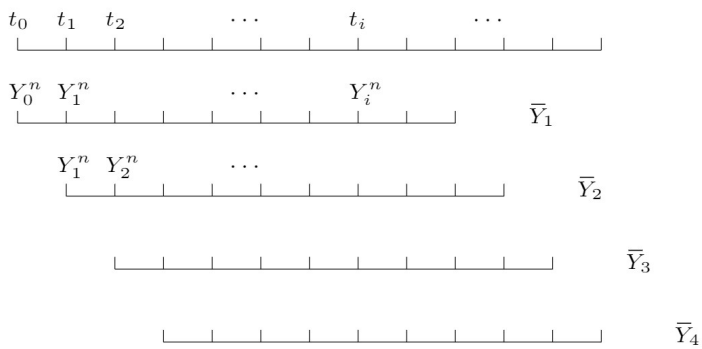
$$\widehat{c}_i^n \equiv \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n-l_n+1} \left(\overline{Y}_{i+h}^n \otimes \overline{Y}_{i+h}^n \mathbb{1}_{\{\|\overline{Y}_{i+h}^n\| \le \nu_n\}} - \widehat{Y}_{i+h}^n \right) \tag{3.3}$$

This offset term $\widehat{Y}_i^n$ is needed to achieve the optimal convergence rate in CLT. Yet a disadvantage is that $\widehat{c}_i^n$ can not be guaranteed to be positive semidefinite in finite samples. A fix is to project $\widehat{c}_i^n$ onto the convex cone $\mathcal{S}_d^+$ in Frobenius norm. Suppose $\widehat{c}_i^n = Q\Lambda Q^{\mathrm{T}}$ is the eigenvalue factorization, the positive semidefinite projection is $\widehat{c}_i^{n} = Q\Lambda_+ Q^{\mathrm{T}}$ where $\Lambda_+^{ij} = (\Lambda^{jj} \vee 0)\mathbb{1}_{\{i=j\}}$.
Stimulated by [14], we design another estimator of stochastic volatility matrix to fulfill the positive semidefinite requirement:

$$\widetilde{c}_i^n \equiv \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n-l_n+1} \overline{Y}_{i+h}^n \otimes \overline{Y}_{i+h}^n \mathbb{1}_{\{\|\overline{Y}_{i+h}^n\| \le \nu_n\}} \tag{3.4}$$

In addition to being positive semidefinite which some readers may prioritize, (3.4) leads to easier and more practical bias correction and uncertainty quantification for general functionals of volatility matrix (our end goal). This kind of advantage of (3.4) is pronounced in PCA. Detailed comparison between these two estimators is given in Remark 2.

The estimators of stochastic volatility matrix use local moving averages computed on overlapping windows. Local moving averages on overlapping windows fully utilize data and can lead to the optimal convergence rate. The computation can be done via convolution of the noisy data with a sequence induced by the smoothing kernel φ. Hence (3.3) and (3.4) can be implemented by a fast Fourier transform (FFT) algorithm and they are scalable for very large datasets.



3.2. *Correction of nonlinearity bias.* As an exposition of the necessity of nonlinear bias correction, let first consider the case of $d = 1$ and constant volatility. By Taylor expansion, the estimation error of the plug-in estimator $g(\widehat{c})$ can be decomposed as

$$g(\widehat{c}) - g(c) = \underbrace{\partial g(c)(\widehat{c} - c)}_{\text{limit}} + \underbrace{\frac{1}{2}\partial^2 g(c)(\widehat{c} - c)^2}_{\text{nonlinearity bias}} + \underbrace{O_p(|\widehat{c} - c|^3)}_{\text{asymptotically negligible}}$$

As it turns out (shown in Appendix C, D), the first-order term converges to the limit in CLT. The bias arises inevitably, as long as g is nonlinear (with non-zero Hessian), from the second-order derivative and the quadratic form of estimation error.

In the multivariate case, the bias term is more involved. Let $\widehat{B}(g)_i^n$ be the nonlinear bias correction term to $g(\widehat{c}_i^n)$, we find that $\widehat{B}(g)_i^n$ need to be

$$\widehat{B}(g)_i^n = \frac{1}{2k_n \Delta_n^{1/2}} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\widehat{c}_i^n) \times \Xi\big(\widehat{c}_i^n, \widehat{\gamma}_i^n\big)^{jk,lm} \tag{3.5}$$

where Ξ is a tensor-valued function defined in (4.3), and $\widehat{\gamma}_i^n$ is an estimator of the instantaneous covariance matrix of noise at time $i\Delta_n$:

$$\widehat{\gamma}_i^n \equiv \frac{1}{2m_n} \sum_{h=1}^{m_n} \Delta_{i+h}^n Y \otimes \Delta_{i+h}^n Y \tag{3.6}$$

with $m_n = \lfloor \theta' \Delta_n^{-1/2} \rfloor$ and θ' being positive finite.

The nonlinearity bias correction term $\widetilde{B}(g)_i^n$ to $g(\widetilde{c}_i^n)$ is

$$\widetilde{B}(g)_i^n = \frac{1}{2k_n \Delta_n^{1/2+\delta}} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\widetilde{c}_i^n) \times \Sigma\big(\widetilde{c}_i^n\big)^{jk,lm} \tag{3.7}$$

where δ is introduced in (3.12), Σ is defined in (4.2).

3.3. *Estimators of functionals.*

DEFINITION 1. *Let $N_t^n = \lfloor t/(k_n\Delta_n)\rfloor$, a rate-optimal estimator of (1.1) is defined as*

$$\widehat{S}(g)_t^n \equiv k_n\Delta_n \sum_{i=0}^{N_t^n-1} \left[g(\widehat{c}_{ik_n}^n) - \widehat{B}(g)_{ik_n}^n\right] \times a_t^n$$

$\widehat{c}_h^n$ is defined in (3.3), $\widehat{B}(g)_h^n$ is defined in (3.5), $a_t^n = t/(N_t^n k_n\Delta_n)$ is a finite-sample adjustment.

We choose to compute $\widehat{c}_h^n$'s over disjoint windows for estimating $S(g)_t$. One may choose to compute $\widehat{c}_h^n$'s on overlapping windows instead and define an alternative estimator as $\widehat{S}(g)_t'^n \equiv \Delta_n \sum_{i=0}^{\lfloor t/\Delta_n\rfloor-1}[g(\widehat{c}_i^n) - B(g)_i^n]$. But we advice against this alternative for two reasons. The first one is that $\widehat{S}(g)_t^n$ and $\widehat{S}(g)_t'^n$ has the same asymptotic properties such as convergence rate and limiting distributions, cf. [4, 21]. The second one, more essential for some applications, is that computing $\widehat{S}(g)_t'^n$ is unfortunately impractical in some applications, such as the data analysis using PCA in Section 6. In fact, should $\widehat{S}(g)_t'^n$ be employed, we need to compute more than 100,000 eigenvalue factorizations of 90×90 matrices for data of every week, with more than 850 weeks our data analysis would take too long to finish.

DEFINITION 2. *Let $N_t^n = \lfloor t/(k_n\Delta_n)\rfloor$, an estimator of (1.1) with positive semidefinite plug-ins is defined as*

$$\widetilde{S}(g)_t^n \equiv k_n\Delta_n \sum_{i=0}^{N_t^n-1} \left[g(\widetilde{c}_{ik_n}^n) - \widetilde{B}(g)_{ik_n}^n\right] \times a_t^n$$

where $\widetilde{c}_h^n$ is defined in (3.4), $\widetilde{B}(g)_h^n$ is defined in (3.7), and $a_t^n = t/(N_t^n k_n\Delta_n)$ is an adjustment for the right-edge effect.

Our estimators, given the smoothing kernel φ, entail three tuning parameters.

TABLE 1
Tuning parameters

	description
l_n	length of overlapping windows for local moving averages
k_n	length of disjoint windows for stochastic volatility matrix estimation
ν_n	threshold of jump truncation

A judicially tuning choice is an indispensable if not the most important part of sound estimation and inference. As for k_n, there are three levels of considerations:

level 1: k_n should be chosen to result in consistency and a satisfactory convergence rate;
level 2: k_n should be chosen so that we can establish CLT for inference;
level 3: k_n should be chosen in a way that allows convenient correction of nonlinear bias.

A severe pitfall looms when level 3 is dismissed. If falling prey to this pitfall, researchers have to estimate the volatility of volatility and volatility jumps (both are notoriously hard to estimate) in order to establish CLT [28]. Under the no-noise assumption, this pitfall is documented and avoided by [21]. Recently [5] proposed the multiscale jackknife method and [26] used matrix calculus for practical nonlinear bias correction, and they investigated how

to choose k_n more robustly under the no-noise assumption. Their results greatly facilitate statistical applications using sparsely subsampled data.

When we must confront the noise problem to improve statistical quality, there is one more tuning parameter, namely l_n, to consider. However, when noise is taken into account, the three levels of considerations for k_n are missing so far, and given a legitimate l_n how to adjust v_n is unknown.

We provide an answer to this tuning problem. We find the range of tuning parameters when noise and jump are both present. The benefits include more usable data and better estimation accuracy. Importantly, practical bias correction and uncertainty quantification are possible under our tuning configuration. We will show, with suitable choices of the combination (l_n, k_n, ν_n), our estimators are widely applicable to any $g : \mathcal{S}_d^+ \mapsto \mathbb{R}^r$ that satisfies

$$g \in \mathcal{C}^3(\mathcal{S}) \tag{3.8}$$

where $\mathcal{S} \subset \mathcal{S}_d^+$ is the subspace containing an enlargement of the sample path of c. Precisely, $\mathcal{S} \supset \cup_m \mathcal{S}_m^\epsilon$ for some $\epsilon > 0$, $\mathcal{S}_m^\epsilon = \left\{A \in \mathcal{S}_d^+ : \inf_{M \in \mathcal{S}_m} \|A - M\| \le \epsilon\right\}$ is an enlargement of $\mathcal{S}_m$ which is identified in Assumption A-ν.

3.4. *Tuning parameters for the optimal convergence rate.* A proper combination of the tuning parameters is the requisite for consistency, the existence of CLT, and the optimal convergence rate. To fulfill these objectives for $\widehat{S}(g)_t^n$, we need

$$\begin{cases} l_n = \lfloor \theta \Delta_n^{-1/2} \rfloor & \\ k_n = \lfloor \varrho \Delta_n^{-\kappa} \rfloor & \kappa \in \left(\frac{2}{3} \vee \frac{2+\nu}{4}, \frac{3}{4}\right) \\ \nu_n = \alpha \Delta_n^{\rho} & \rho \in \left[\frac{1}{4} + \frac{1-\kappa}{2-\nu}, \frac{1}{2}\right) \end{cases} \tag{3.9}$$

where $\theta, \varrho, \alpha > 0$ are positive finite, and $\nu \in [0, 1)$ is introduced in Assumption A-ν which dictates the jump intensity.

Here we offer an intuition for (3.9). The reader can skip this part without affecting understanding of the main results in Section 4.

- *l_n influences the convergence rate*
  Under the noise model (2.2), according to the definition (3.2),
  $$\overline{Y}_i^n = \overline{X}_i^n + \overline{\varepsilon}_i^n$$
  where $\overline{X}_i^n$ and $\overline{\varepsilon}_i^n$ are defined analogously to $\overline{Y}_i^n$ (see (3.2)). We can write $\overline{\varepsilon}_i^n = -\psi_n^{-1/2} \sum_{h=0}^{l_n-1} (\varphi_{h+1}^n - \varphi_h^n)\, \varepsilon_{i+h}^n$. By the conditional independence of ε_i^n's, $\overline{\varepsilon}_i^n = O_p(l_n^{-1})$; $\overline{X}_i^n = O_p(\Delta_n^{-1/2})$ by (C.7) in Appendix C.2. By taking $l_n \asymp \Delta_n^{-1/2}$ the orders of $\overline{X}_i^n$ and $\overline{\varepsilon}_i^n$ are equal. It turns out that this choice of local smoothing window will deliver the optimal rate of convergence.
- *k_n dictates bias-correction and the CLT form*
  Here for exposition let's suppose $d = 1$ and X is an continuous Itô semimartingale, then
  $$\widehat{c}_i^n - c_i^n = d_i^n + s_i^n$$
  where
  $$d_i^n = \frac{1}{(k_n - l_n)\Delta_n} \int_{i\Delta_n}^{(i+k_n-l_n+1)\Delta_n} (c_s - c_i^n)\, \mathrm{d}s$$
  $$s_i^n = \frac{1}{(k_n - l_n)\Delta_n} \left[\widehat{c}_i^n (k_n - l_n)\Delta_n - \int_{i\Delta_n}^{(i+k_n-l_n+1)\Delta_n} c_s\, \mathrm{d}s\right]$$

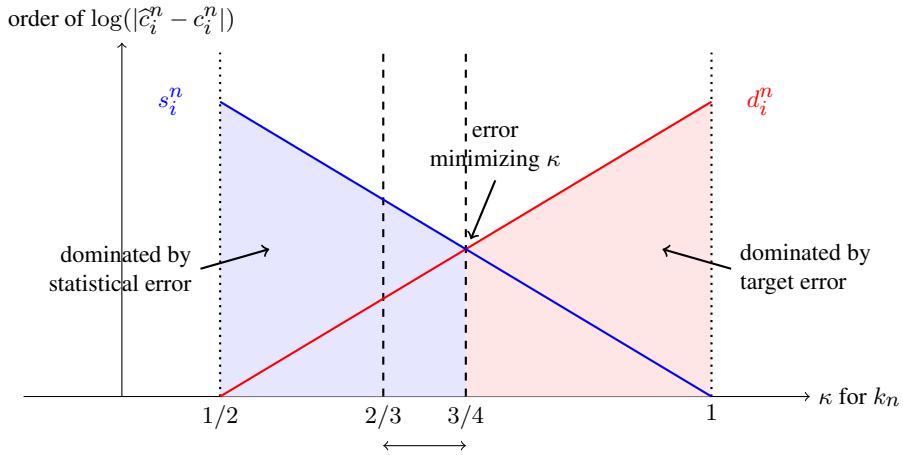


FIG 4. *A schematic plot of tuning k_n*

– We call d_i^n "target error"; $d_i^n = O_p((k_n\Delta_n)^{1/2})$ by (B.2) in Appendix B; its value depends on the temporal evolution of the underlying stochastic volatility.
– We call s_i^n "statistical error"; $s_i^n \approx \frac{1}{(k_n-l_n)\Delta_n}\Delta_n^{1/4}(\chi^n_{i+k_n-l_n+1} - \chi_i^n)$, where χ is another continuous Itô semimartingale; this result is due to (3.8) in [12]; so $s_i^n = O_p((k_n\Delta_n^{1/2})^{-1/2})$.

Balancing the orders of d_i^n and s_i^n by setting $\kappa = 3/4$ brings the error $\widehat{c}_i^n - c_i^n$ down to the minimum order. However, for the purpose of general nonlinear functionals, choosing $\kappa \geq 3/4$ causes the bias to be substantially influenced by d_i^n which depends on volatility of volatility and volatility jump (difficult to estimate and de-bias in applications). Therefore, it is advisable to choose $\kappa < 3/4$, in which case s_i^n dominates and d_i^n vanishes asymptotically, thereby the thorny terms are circumvented. Besides, to achieve successful de-biasing of statistical error and negligibility of higher-order Taylor-expansion terms, we need $\kappa > 2/3$.

When jump is present and its activity index $\nu > 2/3$, the lower-bound requirement on κ increase from $2/3$ to $(2+\nu)/4$. This increase is required for uniform convergence in the spatial localization procedure, which in turn is necessary for functionals with a singularity at the origin.

Section 3.1, 3.2 of [21] give an analogous discussion and their conclusion is $1/3 < \kappa < 1/2$ under the no-noise assumption. With noise and local moving averages, our conclusion is markedly different.

- *ν_n disentangles volatility from jump variation*
According to (C.1) in Appendix C.1, $\|\overline{Y}_i^n\| = O_p(\Delta_n^{1/2})$ if there is no jump over $[i\Delta_n, (i+l_n)\Delta_n]$. By choosing $\rho < 1/2$, the truncation level $\nu_n > \Delta_n^{1/2}$, hence jump truncation keeps the diffusion movements largely intact and discards jumps in a certain sense. To effectively filter out the jumps, the truncation level should be bounded from above and the upper bounds depends on the jump activity index ν.

REMARK 1. If the reader is interested to estimate (1.1) with g satisfying

$$\|\partial_h g(x)\| \leq K_h(1+\|x\|^{r-h}),\ h = 0,1,2,3,\ r \geq 3 \tag{3.10}$$

the requirements on k_n and ν_n can be loosened and become

$$\begin{array}{ll} k_n = \lfloor \varrho\Delta_n^{-\kappa}\rfloor & \kappa \in \left(\frac{2}{3}, \frac{3}{4}\right) \\ \nu_n = \alpha\Delta_n^{\rho} & \rho \in \left[\frac{1}{4} + \frac{1}{4(2-\nu)}, \frac{1}{2}\right) \end{array} \tag{3.11}$$

For generality and wider applicability, we choose to accommodate the functional space (3.8) and retain the requirement (3.9).

3.5. *Tuning parameters for p.s.d. plug-ins.* If readers choose to use $\widetilde{S}(g)_i^n$, the tuning parameters should satisfy

$$\begin{cases} l_n = \lfloor \theta \Delta_n^{-1/2-\delta} \rfloor & \delta \in \left(\frac{1}{10}, \frac{1}{2}\right) \\ k_n = \lfloor \varrho \Delta_n^{-\kappa} \rfloor & \kappa \in \left(\left(\frac{2}{3} + \frac{2\delta}{3}\right) \vee \left(\frac{2+\nu}{4} + \frac{(2-\nu)\delta}{2}\right), \frac{3}{4} + \frac{\delta}{2}\right) \\ \nu_n = \alpha \Delta_n^{\rho} & \rho \in \left[\frac{1}{4} + \frac{\delta}{2} + \frac{1-\kappa}{2-\nu}, \frac{1}{2}\right) \end{cases} \tag{3.12}$$

The key insight is that we can increase l_n to attenuate noise in $\overline{Y}_i^n$ and dispense with the noise offset term $\widehat{Y}_i^n$. However, the positive semidefinite estimator does not possess the optimal convergence rate (see the next section).

4. Asymptotics and Inference.

4.1. *Elements.* Before stating the asymptotic result, we need to introduce some elements in our asymptotic theory. We associate the following quantities with the smoothing kernel φ for $l, m = 0, 1$:

$$\begin{aligned} &\phi_0(s) = \textstyle\int_s^1 \varphi(u)\varphi(u-s)\,\mathrm{d}u, && \phi_1(s) = \textstyle\int_s^1 \varphi'(u)\varphi'(u-s)\,\mathrm{d}u \\ &\Phi_{lm} = \textstyle\int_0^1 \phi_l(s)\phi_m(s)\,\mathrm{d}s, && \Psi_{lm} = \textstyle\int_0^1 s\,\phi_l(s)\phi_m(s)\,\mathrm{d}s \end{aligned} \tag{4.1}$$

We define Σ, Θ, Υ as $\mathbb{R}^{d\times d\times d\times d}$-valued functions, such that for $x, z \in \mathbb{R}^{d\times d}$, $j, k, l, m = 1, \cdots, d$,

$$\begin{aligned} \Sigma(x)^{jk,lm} &= \frac{2\theta\,\Phi_{00}}{\phi_0(0)^2}\left(x^{jl}x^{km} + x^{jm}x^{kl}\right) \\ \Theta(x,z)^{jk,lm} &= \frac{2\Phi_{01}}{\theta\,\phi_0(0)^2}\left(x^{jl}z^{km} + x^{jm}z^{kl} + x^{km}z^{jl} + x^{kl}z^{jm}\right) \\ \Upsilon(z)^{jk,lm} &= \frac{2\Phi_{11}}{\theta^3\,\phi_0(0)^2}\left(z^{jl}z^{km} + z^{jm}z^{kl}\right) \end{aligned} \tag{4.2}$$

where θ is introduced in (3.9). We define Ξ as a tensor-valued function

$$\Xi(x,z) = \Sigma(x) + \Theta(x,z) + \Upsilon(z) \tag{4.3}$$

Now we are ready to describe limiting processes.

DEFINITION 3. *Suppose g satisfies (3.8) or (3.10), $Z_\Xi(g)$ and $Z_\Sigma(g)$ are two stochastic processes defined on an extension of the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ specified in (A.4), such that conditioning on $\mathcal{F}$, $Z_\Xi(g)$ and $Z_\Sigma(g)$ are centered Gaussian processes with conditional covariance matrices being*

$$\begin{aligned} \overline{E}[Z_\Xi(g)Z_\Xi(g)^{\mathrm{T}}|\mathcal{F}] &= V_\Xi(g) \\ \overline{E}[Z_\Sigma(g)Z_\Sigma(g)^{\mathrm{T}}|\mathcal{F}] &= V_\Sigma(g) \end{aligned}$$

where $\overline{E}$ is the conditional expectation operator on the extended probability space and

$$V_\Xi(g)_t = \int_0^t \sum_{j,k,l,m=1}^{d} \partial_{jk}g(c_s)\,\partial_{lm}g(c_s)^{\mathrm{T}}\,\Xi(c_s,\gamma_s)^{jk,lm}\,\mathrm{d}s \tag{4.4}$$

$$V_\Sigma(g)_t = \int_0^t \sum_{j,k,l,m=1}^{d} \partial_{jk}g(c_s)\,\partial_{lm}g(c_s)^{\mathrm{T}}\,\Sigma(c_s)^{jk,lm}\,\mathrm{d}s \tag{4.5}$$

where the process γ is the conditional instantaneous covariance matrix of noise defined in (A.5).

4.2. *Central limit theorems.*

THEOREM 1. *Assume assumptions A-ν, A-γ, and suppose g satisfies (3.8). We select the tuning parameters according to (3.9), then we have the following stable convergence in law of discretized process to a conditional continuous Itô semimartingale on compact subsets of $\mathbb{R}^+$:*

$$\Delta_n^{-1/4}\left[\widehat{S}(g)^n - S(g)\right] \xrightarrow{\mathcal{L}-s(f)} Z_\Xi(g) \tag{4.6}$$

where $S(g)$ is defined in (1.1), $\widehat{S}(g)^n$ is from Definition 1, $Z(g)$ is identified in Definition 3.

Theorem 1 is valid over the functional space (3.8), which is as general as the state-of-the-art literature can achieve. If applications require functionals whose derivatives satisfy a stronger condition (3.10) of polynomial growth, we can put less restrictions on the tuning parameters.

PROPOSITION 1. *Assume Assumptions A-ν, A-γ. Replace the functional space (3.8) with (3.10), replace the tuning conditions (3.9) on k_n, ν_n with (3.11), then the functional stable convergence (4.6) still holds true.*

However, Proposition 1 rules out functionals that involve matrix inversion, it is not applicable to, for instance, estimation of precision matrices and inference of linear regression models. In the rest of this paper, we focus on the results over the general functional space (3.8).
We also establish a limit theorem for $\widetilde{S}(g)^n$ using positive semidefinite plug-ins:

THEOREM 2. *Assume assumptions A-ν, A-γ, and suppose g satisfies (3.8). We select the tuning parameters according to (3.12), then we have the following stable convergence in law of discretized process to a conditional continuous Itô semimartingale on compact subsets of $\mathbb{R}^+$:*

$$\Delta_n^{-(1/4-\delta/2)}\left[\widetilde{S}(g)^n - S(g)\right] \xrightarrow{\mathcal{L}-s(f)} Z_\Sigma(g) \tag{4.7}$$

where δ is specified in (3.12), $Z_\Sigma(g)$ is identified in Definition 3.

The foregoing asymptotic results are stated in terms of converging stochastic processes, which is necessary to express functional stable convergence in law (the strongest mode of convergence one can obtain in our setting). Below is a point-wise formulation which is more relevant for statistical applications. For finite t, and under the conditions of Theorem 1,

$$n^{1/4}\left[\widehat{S}(g)_t^n - S(g)_t\right] \xrightarrow{\mathcal{L}-s} \mathcal{MN}\big(0, \sqrt{t}\, V_\Xi(g)_t\big) \tag{4.8}$$

Moreover, under the conditions of Theorem 2,

$$n^{1/4-\delta/2}\left[\widetilde{S}(g)_t^n - S(g)_t\right] \xrightarrow{\mathcal{L}-s} \mathcal{MN}\big(0, t^{1/2-\delta}\, V_\Sigma(g)_t\big) \tag{4.9}$$

4.3. *Confidence intervals.* The asymptotic variances (4.4) and (4.5) are also functionals of volatility matrix, so we can also estimate the asymptotic variances by the plug-in framework, i.e. plugging in estimates of stochastic volatility matrices (and estimates of instantaneous noise covariance matrices) to the corresponding functional forms.

DEFINITION 4. *For g that satisfies (3.8), we define*

$$\widehat{V}_{\Xi}(g)_t^n \equiv k_n \Delta_n \sum_{i=0}^{N_t^n-1} \sum_{j,k,l,m=1}^{d} \partial_{jk} g(\widehat{c}_{ik_n}^n)\, \partial_{lm} g(\widehat{c}_{ik_n}^n)^{\mathrm{T}}\, \Xi(\widehat{c}_{ik_n}^n, \widehat{\gamma}_{ik_n}^n)^{jk,lm} \tag{4.10}$$

as the estimator for $V_{\Xi}(g)_t$ for which we abide the tuning (3.9).
Furthermore, we define

$$\widetilde{V}_{\Sigma}(g)_t^n \equiv k_n \Delta_n \sum_{i=0}^{N_t^n-1} \sum_{j,k,l,m=1}^{d} \partial_{jk} g(\widetilde{c}_{ik_n}^n)\, \partial_{lm} g(\widetilde{c}_{ik_n}^n)^{\mathrm{T}}\, \Sigma(\widetilde{c}_{ik_n}^n)^{jk,lm} \tag{4.11}$$

as the estimator for $V_{\Sigma}(g)_t$ for which we abide the tuning (3.12).

We have that both $\widehat{V}_{\Xi}(g)^n$ and $\widetilde{V}_{\Sigma}(g)^n$ are consistent:

PROPOSITION 2. *Suppose Assumptions A-ν, A-γ, g satisfies (3.8) and t is finite.*
With the tuning (3.9), we have

$$\left\|\widehat{V}_{\Xi}(g)_t^n - V_{\Xi}(g)_t\right\| = O_p(\Delta_n^{\kappa-1/2})$$

With the tuning (3.12), we have

$$\left\|\widetilde{V}_{\Sigma}(g)_t^n - V_{\Sigma}(g)_t\right\| = O_p(\Delta_n^{\kappa-1/2-\delta})$$

PROOF. The asymptotic variance $V_{\Xi}(g)_t$ is a smooth functional of volatility matrix and instantaneous noise covariance matrix. Its consistency follows from the consistence of the volatility matrix estimator (3.3) and the instantaneous noise covariance matrix estimator (3.6). According to Lemma 5 in Appendix C.2 and (C.21) in Appendix C.4, the error rate of $\widehat{V}_{\Xi}(g)_t^n$ is determined by the estimation error of (3.3). Therefore, the error rate of $\widehat{V}_{\Xi}(g)_t^n$ is the same as the error rate of the volatility functional estimator without bias correction, which is $(k_n\Delta_n^{1/2})^{-1}$, then the proposition follows from (3.9).
The consistency of $\widetilde{V}_{\Sigma}(g)_t^n$ is established by a similar argument. □

Based on Theorem 1, Theorem 2, Proposition 2 and the property of stable convergence, we have the following feasible central limit theorems:

COROLLARY 1. *Suppose g satisfies (3.8) and Assumptions A-ν, A-γ hold. With the tuning (3.9), we have*

$$\left[\Delta_n^{1/2}\, \widehat{V}_{\Xi}(g)_t^n\right]^{-1/2} \left[\widehat{S}(g)_t^n - S(g)_t\right] \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \mathbf{I}_r\big) \tag{4.12}$$

in restriction to the event $\{\omega \in \Omega, \widehat{V}_{\Xi}(g)_t^n$ is positive definite$\}$, where Ω is defined in (A.4).
With the tuning (3.12), we have

$$\left[\Delta_n^{1/2-\delta}\, \widetilde{V}_{\Sigma}(g)_t^n\right]^{-1/2} \left[\widetilde{S}(g)_t^n - S(g)_t\right] \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \mathbf{I}_r\big) \tag{4.13}$$

in restriction to the event $\{\omega \in \Omega, \widetilde{V}_{\Sigma}(g)_t^n$ is positive definite$\}$.

5. Applications.

5.1. *PCA of high-frequency data.* PCA is one of powerful statistical methods that help us obtain information and gain insights from large and complex datasets. It reduces dimensionality, facilitates data visualization, reveals common trends et cetera. For high-frequency data, PCA can be done by inference of spectral structure of stochastic volatility matrix

$$c_t = \left[q_t^1, \cdots, q_t^d\right] \begin{bmatrix} \lambda_t^1 & & \\ & \ddots & \\ & & \lambda_t^d \end{bmatrix} \begin{bmatrix} q_t^{1,\mathrm{T}} \\ \vdots \\ q_t^{d,\mathrm{T}} \end{bmatrix} \tag{5.1}$$

The eigenvalues and eigenvectors can be considered as functionals of the associated stochastic volatility matrix once directions and lengths of the eigenvectors are specified and some regularity conditions hold. As per this functional idea we write $\lambda^k(c_t) = \lambda_t^k$, $q^k(c_t) = q_t^k$. Based on this insight, [4] applied PCA to nonstationary financial data by conducting inference on the realized eigenvalue $\int_0^t \lambda_s^k \,\mathrm{d}s$, realized eigenvector $\int_0^t q_s^k \,\mathrm{d}s$, realized principal component $\int_0^t q_{s-}^k \,\mathrm{d}X_s$. More recently, [27] extends the realized PCA to asynchronously observed high-frequency noisy data. Here we extend [4] to noisy data, and improve the convergence rate of [27] by a considerable margin.
Suppose for $t \in [0, T]$,

$$\lambda_t^{r_0+1} = \cdots = \lambda_t^{r_1} > \lambda_t^{r_1+1} = \cdots = \lambda_t^{r_2} > \cdots\cdots > \lambda_t^{r_{K-1}+1} = \cdots = \lambda_t^{r_K} \tag{5.2}$$

where $r_0 = 0$, $r_K = d$. Let $\mathcal{K}_k = \{r_{k-1}+1, \cdots, r_k\}$ for $k = 1, \cdots, K$. Define the following functions of volatility matrix

$$\lambda(c_t) = \Big(\frac{1}{r_1 - r_0} \sum_{r \in \mathcal{K}_1} \lambda_t^r, \cdots, \frac{1}{r_K - r_{K-1}} \sum_{r \in \mathcal{K}_K} \lambda_t^r\Big)^{\mathrm{T}}$$

If $r_k = r_{k-1} + 1$, λ_t^k is called a simple eigenvalue, otherwise a repeated eigenvalue. If λ_t^k is a repeated eigenvalue, it is not differentiable with respect to c_t, and establishing its asymptotic distribution would be problematic. Nevertheless, λ defined above is a differentiable function regardless of singleton or multicity (see Remark 3 in Appendix E), so we can do nonlinear bias correction and derive its CLT.
To do PCA, we need to estimate

$$S(\lambda)_t = \int_0^t \lambda(c_s) \,\mathrm{d}s$$

$$S(q^k)_t = \int_0^t q^k(c_s) \,\mathrm{d}s, \quad k = 1 \cdots, d$$

Given $\widetilde{c}_i^n$ defined in (3.4), we compute its spectral factorization:

$$\widetilde{c}_i^n = \left[\widetilde{q}_i^{1,n}, \cdots, \widetilde{q}_i^{d,n}\right] \begin{bmatrix} \widetilde{\lambda}_i^{1,n} & & \\ & \ddots & \\ & & \widetilde{\lambda}_i^{d,n} \end{bmatrix} \left[\widetilde{q}_i^{1,n}, \cdots, \widetilde{q}_i^{d,n}\right]^{\mathrm{T}}$$

Define $\overline{\lambda}_i^{k,n} = (r_k - r_{k-1})^{-1} \sum_{r \in \mathcal{K}_k} \widetilde{\lambda}_i^{r,n}$, our estimator of $S(\lambda)_t$ is

$$\widetilde{S}(\lambda)_t^n = \left[\widetilde{S}(\lambda)_t^{1,n}, \cdots, \widetilde{S}(\lambda)_t^{K,n}\right]^{\mathrm{T}}$$

$$\widetilde{S}(\lambda)_t^{k,n} \equiv k_n \Delta_n \sum_{h=0}^{N_t^n - 1} \left[1 - \frac{2\theta\Phi_{00}}{\phi_0(0)^2 \, k_n \, \Delta_n^{1/2+\delta}} \sum_{v \notin \mathcal{K}_k} \frac{\widetilde{\lambda}_{hk_n}^{v,n}}{\overline{\lambda}_{hk_n}^{k,n} - \widetilde{\lambda}_{hk_n}^{v,n}}\right] \overline{\lambda}_{hk_n}^{k,n} \tag{5.3}$$

where θ, Φ_{00}, $\phi_0(0)$ are defined in (3.9), (4.1). Note that $\widetilde{S}(\lambda)_t^{k,n}$ is a Riemann sum of simple eigenvalues or averages of repeated eigenvalues, with a multiplicative bias-correction factor. We have an expression for this multiplicative factor in terms of eigenvalue estimates. Since we do not have to evaluate Hessian of the functional, (5.3) is very advantageous for applications.

This advantage is also carried over to eigenvectors. We derive a closed-form expression for the nonlinear bias correction to the eigenvectors and define the estimator of $S(q^k)_t$ to be

$$\widetilde{S}(q^k)_t^n \equiv k_n \Delta_n \sum_{h=0}^{N_t^n-1} \left[1 + \frac{\theta \Phi_{00}}{\phi_0(0)^2\, k_n\, \Delta_n^{1/2+\delta}} \sum_{v \neq k} \frac{\widetilde{\lambda}_{hk_n}^{k,n} \widetilde{\lambda}_{hk_n}^{v,n}}{\left(\widetilde{\lambda}_{hk_n}^{k,n} - \widetilde{\lambda}_{hk_n}^{v,n}\right)^2} \right] \widetilde{q}_{hk_n}^{k,n} \tag{5.4}$$

PROPOSITION 3. *Assume Assumption A-ν, A-γ, and suppose (5.2). We select the bandwidths l_n, k_n and the truncation threshold ν_n according to (3.12). We have the following functional stable convergence in law:*

$$\Delta_n^{-1/4+\delta/2} \left[\widetilde{S}(\lambda)^n - S(\lambda) \right] \xrightarrow{\mathcal{L}-s(f)} Z(\lambda)$$

where $Z(\lambda)$ is a process defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, such that conditioning on $\mathcal{F}$ it is a continuous centered Gaussian martingale with variance

$$\overline{E}[Z(\lambda)Z(\lambda)^{\mathrm{T}}|\mathcal{F}] = V(\lambda)$$

where $\overline{E}$ is the conditional expectation operator on the extended probability space and

$$V(\lambda)_t = \frac{4\theta\Phi_{00}}{\phi_0(0)^2} \begin{bmatrix} (r_1 - r_0)^{-1} \int_0^t (\lambda_s^{r_1})^2 \,\mathrm{d}s & & \\ & \ddots & \\ & & (r_K - r_{K-1})^{-1} \int_0^t (\lambda^{r_K})^2 \,\mathrm{d}s \end{bmatrix} \tag{5.5}$$

If λ_t^k is a simple eigenvalue, the corresponding eigenvector q_t^k is differentiable with respect to c_t (see Lemma 10 in Appendix E). In this case we can derive an inferential theory for the integrated eigenvector estimator (5.4).

PROPOSITION 4. *Assume Assumption A-ν, A-γ, and suppose λ_t^k is a simple eigenvalue of c_t, $\forall t \in [0,T]$. We control the bandwidths l_n, k_n and the truncation threshold ν_n according to (3.12). We have the following functional stable convergence in law:*

$$\Delta_n^{-1/4+\delta/2} \left[\widetilde{S}(q^k)^n - S(q^k) \right] \xrightarrow{\mathcal{L}-s(f)} Z(q^k)$$

where $Z(q^k)$ is a process defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, such that conditioning on $\mathcal{F}$ it is a continuous centered Gaussian martingale with variance

$$\overline{E}[Z(q^k)Z(q^k)^{\mathrm{T}}|\mathcal{F}] = V(q^k)$$

where $\overline{E}$ is the conditional expectation operator on the extended probability space and

$$V(q^k)_t = \frac{2\theta\Phi_{00}}{\phi_0(0)^2} \int_0^t \sum_{v \neq k} \frac{\lambda_s^k \lambda_s^v}{(\lambda_s^k - \lambda_s^v)^2}\, q_s^v\, q_s^{v,\mathrm{T}} \,\mathrm{d}s \tag{5.6}$$

REMARK 2. Our asymptotic theory for PCA is based on the general result of $\widetilde{S}(g)_t^n$. We compare $\widehat{S}(g)_t^n$ and $\widetilde{S}(g)_t^n$ below in view of the asymptotics and practicality of PCA.

- Convergence rates: $\widehat{S}(g)_t^n$ attains the optimal convergence rate $n^{1/4}$; the convergence rate of $\widetilde{S}(g)_t^n$ is $n^{1/4-\delta/2}$ which is less than $n^{1/5}$ according to (3.12).
- Dealing with negative eigenvalue estimate(s): If $\widehat{c}_i^n$ is not positive semidefinite, we can project it to $\mathcal{S}_+^d$. This is equivalent to setting negative eigenvalue estimate(s) to zero. $\widetilde{S}(g)_t^n$ is positive semidefinite by definition and does not suffer this problem.
- Bias correction and uncertainty quantification: For $\widehat{S}(g)_t^n$ we need to evaluate $\partial_{jk}g$ and $\partial^2_{jk,lm}g$ for $j,k,l,m=1,\cdots,d$ in order to compute bias correction $\widehat{B}(g)_i^n$ and asymptotic variance $\widehat{V}_\Xi(g)_t^n$; this can be quite difficult in applications and motivated [5] to propose the multiscale jackknife method. For $\widetilde{S}(g)_t^n$, we can use estimates $\widetilde{\lambda}_i^{k,n}$ and $\widetilde{q}_i^{k,n}$ to compute $\widetilde{B}(g)_i^n$ and $\widetilde{V}_\Sigma(g)_t^n$. Evident from (5.3), (5.4), (5.5), (5.6), they are straightforward to compute without much computational or memory overhead.

We apply Proposition 3 and Proposition 4 to conduct PCA on TAQ data in Section 6. We choose PCA based on $\widetilde{S}(g)_t^n$ because it scales better for the large TAQ data.

5.2. *More examples.*

EXAMPLE 2 (Uncertainty quantification). Asymptotic variances are often functionals of volatility. In the univariate setting, the so-called "quarticity" defined as $\int_0^t c_s^2\,\mathrm{d}s$ appears in the asymptotic variances of many estimators of integrated volatility. The multivariate counterpart involves $\int_0^t c_s^{jl}c_s^{km}+c_s^{jm}c_s^{kl}\,\mathrm{d}s$ in [21] and $\Xi(c_s,\gamma_s)$ in (4.4). For this reason, the volatility functional estimator effectively facilitates uncertainty quantification for various volatility-related estimators.

EXAMPLE 3 (Laplace transform). [1] put forward an estimator of the realized Laplace transform of volatility defined as

$$\int_0^t e^{iwc_s}\,\mathrm{d}s$$

This transform can be viewed as the characteristic function of volatility under the occupation measure. By matching the the moments of realized Laplace transform with those induced by a model, we can estimate model parameter(s) or test the model. An open question noted by [1] is the estimation of realized Laplace transform using noisy data. By the nonparametric estimation of volatility path in the first stage and the Riemann summation of bias-corrected functional plug-ins in the second stage, this paper presents a rate-optimal solution to the open question.

EXAMPLE 4 (Generalized method of moments (GMM)). [2] proposed the generalized method of integrated moments for financial high-frequency data. In estimating an option pricing model, one observes the process $Z_t=(t,X_t,r_t,d_t)$ where X_t is the price of the underlying observed without any noise, r_t is the short-term interest rate, d_t is the dividend yield. One model of the arbitrage-free option price under the risk-neutral probability measure is

$$\beta_t=f(Z_t,c_t;\theta^*)$$

where f is deterministic, θ^* is the true model parameter. The observed option price is often modeled as

$$O_{i\Delta_n}=\beta_{i\Delta_n}+\epsilon_i$$

where ϵ_i is pricing error and $\mathbb{E}(\epsilon_i) = 0$. Let $g(Z_t, c_t; \theta) = \mathbb{E}[O_t - f(Z_t, c_t; \theta)]$, then we have the following integrated moment condition:

$$\int_0^t g(Z_s, c_s; \theta^*)\, \mathrm{d}s = 0$$

Utilizing noisy observations at higher frequencies, this paper provides a means to compute a bias-corrected sample moment function of GMM.

EXAMPLE 5 (Linear regression). In the practice of linear factor models and financial hedging, one faces the tasks of computing the factor loadings and the hedge ratios. These tasks can be formulated as the estimation of the coefficient β in the time-series linear regression model

$$Z_t^c = \beta^{\mathrm{T}} S_t^c + R_t$$

where

$$\begin{cases} S_t = S_0 + \int_0^t b_u^S \,\mathrm{d}u + \int_0^t \sigma_u^S \,\mathrm{d}W_u^S + J_t^S \\ Z_t = Z_0 + \int_0^t b_u^Z \,\mathrm{d}u + \beta^{\mathrm{T}} \int_0^t \sigma_u^S \,\mathrm{d}W_u^S + \int_0^t \sigma_u^R \,\mathrm{d}W_u^R + J_t^Z \end{cases}$$

$\langle W^S, W^R \rangle = 0$, $S_t \in \mathbb{R}^{d-1}$, $Z_t \in \mathbb{R}$, and S^c, Z^c are the continuous parts of the Itô semi-martingales.
Let $X = (S^{\mathrm{T}}, Z)^{\mathrm{T}}$, we can write $X_t = X_0 + \int_0^t b_u \,\mathrm{d}u + \int_0^t \sigma_u \,\mathrm{d}W_u + J_t$ where $b = (b^{S,T}, b^Z)^{\mathrm{T}}$, $W = (W^{S,\mathrm{T}}, W^R)^{\mathrm{T}}$, $J = (J^{S,\mathrm{T}}, J^Z)^{\mathrm{T}}$ and

$$\sigma = \begin{bmatrix} \sigma^S & 0 \\ \beta^{\mathrm{T}} \sigma^S & \sigma^R \end{bmatrix}$$

so

$$c = \sigma\sigma^{\mathrm{T}} = \begin{bmatrix} \sigma^S \sigma^{S,\mathrm{T}} & \sigma^S \sigma^{S,\mathrm{T}} \beta \\ \beta^{\mathrm{T}} \sigma^S \sigma^{S,\mathrm{T}} & \beta^{\mathrm{T}} \sigma^S \sigma^{S,\mathrm{T}} \beta + (\sigma^R)^2 \end{bmatrix} := \begin{bmatrix} c^{SS} & c^{SZ} \\ c^{ZS} & c^{ZZ} \end{bmatrix}$$

hence by letting $g(c) = c^{SS,-1} c^{SZ}$, we have $\beta = t^{-1} S(g)_t$. [3] proposed a method under the no-noise assumption. When the observations contain noise, the methodology of this paper can extend the method of [3] to wider applicability.

6. Empirical analysis of TAQ data. We analyze the transaction data of S&P 100 constituents. Each stock is sampled every second during 09:35 - 15:55 EST, so we have 22,800 observations for each stock on each business day. We focus on data from 09/10/2003 to 12/31/2019. During this period there were 4106 business days in 852 weeks. By focusing on the time window 09:35 - 15:55 EST, we remove overnight returns which contain big jumps and returns incurred in the first and last five minutes of trading hours which are relatively volatile. 10 illiquid assets are excluded from our statistical analysis.
Instantaneous volatility matrix and its eigenvalues and eigenvectors were estimated for every business day. The realized eigenvalues and realized eigenvectors were computed weekly by aggregating the instantaneous eigenvalues and instantaneous eigenvectors.
The temporal evolution of four leading eigenvalue estimates along with confidence intervals are plotted in Figure 5. It shows $\lambda_q / \sum_j \lambda_j$'s which are the proportions of total variation that can be explained by the corresponding principal components. The first eigenvalues are conspicuously separated from the others and indicate the first principal component explains more than 60% of the cross-section variation for the majority of time. Moreover, it is evident from this dataset that the first four leading eigenvalues were simple rather than repeated.

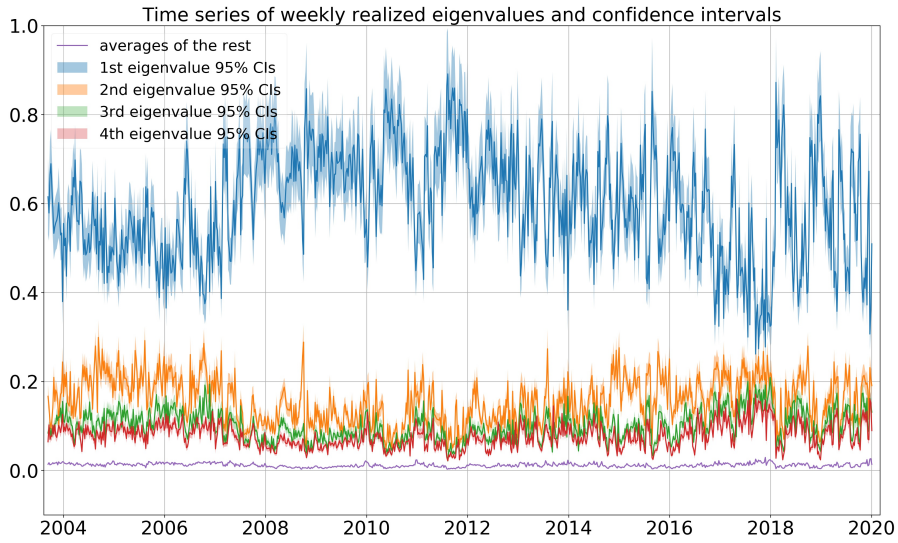


FIG 5. *Weekly integrated eigenvalues and the confidence intervals*

The estimates of eigenvalues tell us that a low-dimensional model can largely explain the systematic comovement of these stocks.

It is reasonable to interpret the first principal component as a market indicator on how the economy performs in general. As all the major corporations and their stock values are moving with the overall economy, the eigenvector corresponding to the first principal component is expected to comprise only positive entries.

An interesting question is how the first principal component is compared with S&P 100 index. [27] calculated the accumulated returns from 2007 to 2017 of the portfolio based on the eigenvector q^1 associated with the first principal component. [27] normalized the eigenvector so that $\sum_j^d q^{1,j} = 1$ and used $q^{1,j}$'s as portfolio weights so that the resultant portfolio is self-financing. A curious finding of [27] is that this portfolio outperformed the market index hence beat the passive investment. Here the author would like to test this portfolio using pre-averaging-based PCA, over a longer time span.

The principal components are $\int_0^t q^r(c_{s-})\,\mathrm{d}X_s$, $r = 1, 2, \cdots$, where $X_t^j = \log(S_t^j)$ and S_t is the vector of asset prices at time t. However, one obviously can not transact in logarithmic currency, so a pending question is how to find a portfolio that is analogous to the first principal component and requires trading on the currency's original scale.

[27] proposed the following construction. Let

$$R_{t,\Delta} = \left[(S_{t+\Delta}^1 - S_t^1)/S_t^1, \cdots, (S_{t+\Delta}^d - S_t^d)/S_t^d\right]^{\mathrm{T}}$$

be the vector of returns in percentage. When $\|R_{t,\Delta}\|$ is small,

$$\log\left(1 + q^{1,\mathrm{T}} R_{t,\Delta}\right) \approx q^{1,\mathrm{T}} R_{t,\Delta},$$

and

$$q^{1,j}\frac{S_{t+\Delta}^j - S_t^j}{S_t^j} \approx q^{1,j}\log\left(1 + \frac{S_{t+\Delta}^j - S_t^j}{S_t^j}\right) = q^{1,j}\log\left(\frac{S_{t+\Delta}^j}{S_t^j}\right),$$

so

$$q^{1,\mathrm{T}}(X_{t+\Delta} - X_t) \approx \log\big(1 + q^{1,\mathrm{T}} R_{t,\Delta}\big).$$

Therefore, we consider $\sum_t \log(1 + q^{1,\mathrm{T}} R_{t,\Delta})$ as an approximation to the first realized principal component. It is also the accumulated log return if there is no transaction cost and $\sum_j^d q_{1,j} = 1$. In this case, $\exp\big\{\sum_t \log(1 + q^{1,\mathrm{T}} R_{t,\Delta})\big\} - 1 = \prod_t (1 + q^{1,\mathrm{T}} R_{t,\Delta}) - 1$ is the accumulated return of this self-financing portfolio in the absence of transaction cost.

Since the volatility matrix and the eigenvector q^1 are updated on every business day, the portfolio can be re-weighted daily. One caveat is that the numerical values of eigenvector estimates are relative less stable. Roughly speaking, the eigenvector viewed as a function of the corresponding positive semidefinite matrix is less smooth than the eigenvalue. Due to this reason, heuristically, an error in the covariance matrix estimation tends to result in a larger error in the eigenvector estimate, as compared to the eigenvalue counterpart. To stabilize eigenvector estimates, it is advisable to compute moving averages of the initial eigenvector estimates over a window of six business days. As a result we re-weight the portfolio on all the business days from 09/17/2019 to 12/31/2019.

Before the portfolio re-allocation according to the new weights on the next business day, one can either clear positions at the end of trading hours or let positions stand overnight. In the second scenario, the portfolio bears the overnight risk premium and absorbs overnight returns.

The accumulated log returns of the portfolio that mimics the first principal component are plotted in Figure 6. The portfolio returns were computed using daily prices of S&P 100 constituents which can be obtained from the Compustat database. As we can see, return morphologies of the portfolio that mimics the first principal component resemble that of S&P 100 index. If the portfolio was cleared at the end of trading hours, it slightly underperformed the index. Interestingly enough, if the portfolio that stood overnight, it manifestly outperformed the index. It is evident that this portfolio and the factor corresponding to the first principal component earned a significant risk premium during the overnight periods, and implies a difference in the intraday and overnight risk-reward patterns. This finding is consistent with the empirical analysis of [27], though we use a method that is more accurate and analyze data over a longer time span.

7. Conclusion. In this paper, we proposed statistical methods for general nonlinear functionals of volatility matrix, and developed statistical theory to guide bias correction, quantify statistical uncertainty and perform statistical inference. Because the methods can utilize data that contains microstructure noise and jumps, and because volatility matrix functionals form a unifying framework that accommodates many applications, this paper provides a viable answer to the open problem of noise and extends numerous statistical methods to much wider applicability. This brings significant statistical gains, such as principled approaches to take advantage of much bigger datasets and improved statistical accuracy.

The tuning choice is an essential component in the inference of volatility matrix functionals, and it has been the main motivation of several recent works under the no-noise assumption. The importance of tuning must not be neglected neither in the noisy setting. We derived tuning ranges for pre-averaging and jump truncation. We would like to stress that the tuning parameters need to be quite different from those under the no-noise assumption.

Asymptotic analysis shows there are non-negligible higher order biases beyond linear terms. We can effectively debias our estimation by quadratic approximation. Given the tuning choice and bias correction, we established consistency, convergence rates, and asymptotic normality by functional stable convergence of processes. We derived stable CLT for both the rate-optimal method and the method that uses positive semidefinite plug-ins. The asymptotic covariance matrices can be consistently estimated, which paves the way for feasible statistical inference.

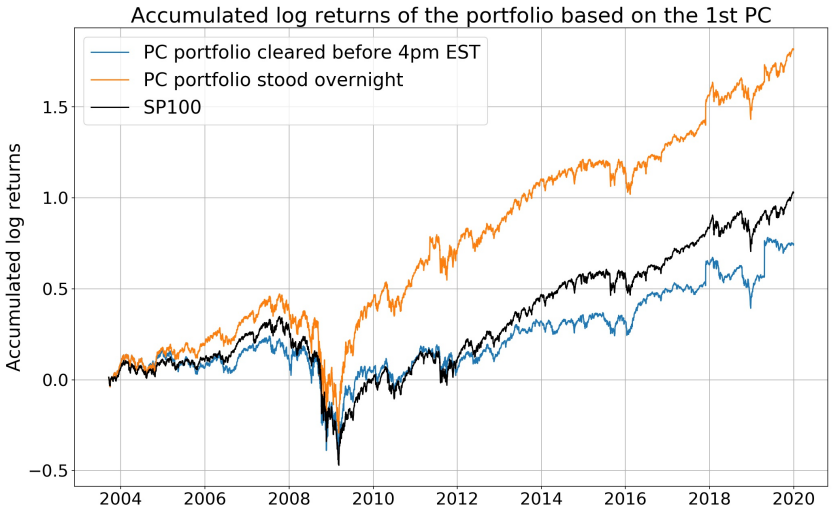


FIG 6. *Accumulated log returns of the portfolio based on the first principal component*

Based on the general results, we then put forward the method and theory of PCA to analyze noisy high-frequency financial data. Particularly we gave a demonstration on the TAQ dataset and analyzed the temporal structures of eigenvalues and the first principal component of S&P 100 stocks. Our PCA method is not only statistically sound in the presence of noise and jump, but also computationally practical for large datasets. PCA is one example among many possible applications of this paper. We demonstrate some success in pushing the inferential framework for general functionals into a new frontier of potentials and possibilities, and lending the effort to enable applications to employ noisy high-frequency data where exciting new discoveries await.

Acknowledgements. The author is indebted to Jean Jacod and Per A. Mykland for encouragement, detailed discussion and comments. This work also benefits from discussions with Yoann Potiron, Viktor Todorov, Dacheng Xiu.
The author was supported by NSF grant DMS 17-13129 and 2019 Stevanovich Fellowship from the University of Chicago.

APPENDIX A: ASSUMPTIONS

We present formally the model specification and assumptions in this appendix.
The pure jump process J in X is modeled by

$$J_t = \int_{(0,t]\times E} \delta(s,x)\, \mathfrak{p}(\mathrm{d}s,\, \mathrm{d}x) \tag{A.1}$$

where δ is a $\mathbb{R}^d$-valued predictable function on $\mathbb{R}^+ \times E$, E is a Polish space, $\mathfrak{p}$ is a Poisson random measure with compensator $\mathfrak{q}(\,\mathrm{d}u,\, \mathrm{d}x) = \mathrm{d}u \otimes \lambda(\,\mathrm{d}x)$, λ is a σ-finite measure on E and has no atom.

The stochastic volatility matrix is positive semidefinite almost surely and of the form

$$c_t = c_0 + \int_0^t b_s^{(c)}\,\mathrm{d}s + \int_0^t \sigma_s^{(c)}\,\mathrm{d}W_s + J_t^{(c)} \tag{A.2}$$

where $b^{(c)} \in \mathbb{R}^{d\times d}$, $\sigma^{(c)} \in \mathbb{R}^{d\times d\times d'}$, $b^{(c)}$ and $\sigma^{(c)}$ are adapted and càdlàg; $J^{(c)}$ is the volatility jump, which is modeled by

$$\begin{aligned} J_t^{(c)} = \delta^{(c)}(s,x)\mathbb{1}_{\{\|\delta^{(c)}(s,x)\|\le 1\}}\,(\mathfrak{p}-\mathfrak{q})(\mathrm{d}s,\,\mathrm{d}x) \\ + \int_{(0,t]\times E} \delta^{(c)}(s,x)\mathbb{1}_{\{\|\delta^{(c)}(s,x)\|>1\}}\,\mathfrak{p}(\mathrm{d}s,\,\mathrm{d}x) \end{aligned} \tag{A.3}$$

where $\delta^{(c)}$ is a $\mathbb{R}^{d\times d}$-valued predictable function on $\mathbb{R}^+\times E$.
To guarantee that c_t is positive semidefinite, one needs to impose additional parametric or semiparametric restrictions in concrete applications. The results in this paper hold true for any process c that is positive semidefinite and satisfies (A.2).
The regularization on drift, diffusion coefficient and jump activity is given below.

ASSUMPTION A-ν (regularity). *The drift process b has $\frac{1}{2}$-Hölder sample path in the mean-square sense, i.e., $\forall t, s \ge 0$,*

$$E\Big(\sup_{u\in[0,s]} \|b_{t+u}-b_t\| \big| \mathcal{F}_t^{(0)}\Big) \le K s^{1/2},\ a.s.;$$

c is of the form (A.2), there is a sequence of triples $(\tau_m, \mathcal{S}_m, \Gamma_m)$, where τ_m is a stopping time and $\tau_m \nearrow \infty$; $\mathcal{S}_m \subset \mathcal{S}_d^+$ is convex, compact such that

$$t \in [0,\tau_m] \Rightarrow c_t \in \mathcal{S}_m;$$

Γ_m is a sequence of bounded λ-integrable functions on E, such that

$$t\in[0,\tau_m] \Longrightarrow \begin{cases} \|b_t\| + \|\sigma_t\| + \|b_t^{(c)}\| + \|\sigma_t^{(c)}\| \le m \\ \|\delta(t,x)\|^\nu \wedge 1 \le \Gamma_m(x),\ \nu\in[0,1) \\ \|\delta^{(c)}(t,x)\|^2 \wedge 1 \le \Gamma_m(x) \end{cases}$$

Let $\big(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)}), \mathbb{P}^{(0)}\big)$ be a filtered probability space with respect to which X and c are $(\mathcal{F}_t^{(0)})$-adapted; let $\big(\Omega^{(1)}, \mathcal{F}^{(1)}, (\mathcal{F}_t^{(1)}), \mathbb{P}^{(1)}\big)$ be another filtered probability space accommodating Y; $\forall t\ge 0$, $\forall A\in\mathcal{F}^{(0)}$, let $Q_t(A,\cdot)$ be a conditional probability measure on $\big(\Omega^{(1)}, \mathcal{F}^{(1)}\big)$.
All the stochastic dynamics above can be described on the filtered extension $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, where

$$\begin{cases} \Omega = \Omega^{(0)}\times\Omega^{(1)} \\ \mathcal{F} = \mathcal{F}^{(0)}\otimes\mathcal{F}^{(1)} \\ \mathcal{F}_t = \bigcap_{s>t}\big(\mathcal{F}_s^{(0)}\otimes\mathcal{F}_s^{(1)}\big), \qquad \widetilde{\mathcal{F}}_t = \bigcap_{s>t}\big(\mathcal{F}^{(0)}\otimes\mathcal{F}_s^{(1)}\big) \\ \mathbb{P}\left(A\times\mathrm{d}\omega\right) = \mathbb{P}^{(0)}(A)\cdot\otimes_{t\ge 0} Q_t(A,\,\mathrm{d}\omega),\ \forall A\in\mathcal{F}^{(0)} \end{cases} \tag{A.4}$$

Let

$$\epsilon_t(\omega) = Y_t(\omega) - X_t$$

and the conditional covariance process of noise can be written as

$$\gamma_t = \int_{\Omega^{(1)}} \epsilon_t(\omega)\,\epsilon_t(\omega)^{\mathrm{T}}\,Q_t(\cdot,\mathrm{d}\omega) \tag{A.5}$$

ASSUMPTION A-γ (noise). $\forall t \in \mathbb{R}^+$,

$$\int_{\Omega^{(1)}} \epsilon_t(\omega)\, Q_t(\cdot,\, \mathrm{d}\omega) = 0$$

$\forall t \neq s$, $\forall A \in \mathcal{F}^{(0)}_{s\wedge t}$

$$\int_{\Omega^{(1)}\times\Omega^{(1)}} \epsilon_t(\omega)\, \epsilon_s(\omega)^{\mathrm{T}}\, Q_t(A,\, \mathrm{d}\omega)\, Q_s(A,\, \mathrm{d}\omega) = 0$$

furthermore,

$$\gamma_t = \gamma_0 + \int_0^t b_s^{(r)}\, \mathrm{d}s + \int_0^t \sigma_s^{(r)}\, \mathrm{d}W_s + \int_{(0,t]\times E} \delta^{(r)}(s,x)\, \mathfrak{p}(\mathrm{d}s,\, \mathrm{d}x)$$

for the same τ_m, Γ_m *in assumption A-ν and* $\forall q \leq 8$

$$t \in [0,\tau_m] \Longrightarrow \begin{cases} \|b_t^{(r)}\| + \|\sigma_t^{(r)}\| + \mathbb{E}_{\Omega^{(1)}}\|\epsilon_t(\omega)\|^q \leq m \\ \|\delta^{(r)}(t,x)\|^2 \wedge 1 \leq \Gamma_m(x) \end{cases}$$

APPENDIX B: THEORETICAL PRELIMINARIES FOR PROOFS

We use the notation and model described in Section 2 and Appendix A in the main article. In the sequel, the constant K changes across lines but remains finite; K_q is a constant depending on q; $a_n \asymp b_n$ means both a_n/b_n and b_n/a_n are bounded for large n; $\mathbb{E}(\cdot)$ denotes the expectation operator on $(\Omega^{(0)}, \mathcal{F}^{(0)})$ or $(\Omega, \mathcal{F})$; $E(\cdot|\mathcal{H})$ denotes the conditional expectation operator, with $\mathcal{H}$ being $\mathcal{F}_t^{(0)}$, $\mathcal{F}_t^{(1)}$, $\mathcal{F}_t$, $\widetilde{\mathcal{F}}_t$; $E_i^n(\cdot)$ denotes $E(\cdot|\mathcal{F}_i^n)$. Six useful results are stated below.

I. By a localization argument from section 4.4.1 in [33], without loss of generality we can assume $\exists$ a constant K, a bounded λ-integrable function Γ on E, a convex compact subset $\mathcal{S} \in \mathcal{S}_d^+$ and $\epsilon > 0$, $g \in \mathcal{C}^3(\mathcal{S}^\epsilon)$ where $\mathcal{S}^\epsilon$ denotes the ϵ-enlargement of $\mathcal{S}$ (see (3.8) in the main article), such that

$$\text{(B.1)} \qquad \begin{cases} \|b\| + \|\sigma\| + \|b^{(c)}\| + \|\sigma^{(c)}\| + \|b_t^{(r)}\| + \|\sigma_t^{(r)}\| + \mathbb{E}\|\epsilon_t\|^q \leq K,\ \forall q \leq 8 \\ \|\delta(t,x)\|^\nu \wedge 1 \leq \Gamma(x),\ \nu \in [0,1) \\ \|\delta^{(c)}(t,x)\|^2 \wedge 1 + \|\delta^{(r)}(t,x)\|^2 \wedge 1 \leq \Gamma(x) \\ c \in \mathcal{S} \end{cases}$$

II. Define a continuous Itô semimartingale with parameters identical to those of (2.1) in the main article

$$X_t' = X_0 + \int_0^t b_s\, \mathrm{d}s + \int_0^t \sigma_s\, \mathrm{d}W_s$$

Let $Y^* = Y - X + X'$. Based on (3.2) in the main article, define

$$\widehat{c}_i^{*n} = \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n - l_n + 1} \left(\overline{Y}_{i+h}^{*n} \cdot \overline{Y}_{i+h}^{*n,\mathrm{T}} - \widehat{Y}_{i+h}^{*n} \right)$$

$$\widetilde{c}_i^{*n} = \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n - l_n + 1} \overline{Y}_{i+h}^{*n} \cdot \overline{Y}_{i+h}^{*n,\mathrm{T}}$$

In the upcoming derivation, $\|\widehat{c}_i^n - \widehat{c}_i^{*n}\|$ is tightly bounded provided ν_n is properly chosen. The focus then will be shifted from $\widehat{c}_i^n$ to $\widehat{c}_i^{*n}$. Ditto for $\widetilde{c}_i^n$ and $\widetilde{c}_i^{*n}$. The volatility "estimators" calculated on continuous sample paths are more tractable.

III. By estimates of Itô semimartingale increments, $\forall$ finite stopping time τ,

$$\text{(B.2)}\qquad \begin{cases} \left\|E\left(X'_{\tau+s}-X'_{\tau}|\mathcal{F}_{\tau}^{(0)}\right)\right\|+ \\ \quad \left\|E\left(c_{\tau+s}-c_{\tau}|\mathcal{F}_{\tau}^{(0)}\right)\right\|+\left\|E\left(\gamma_{\tau+s}-\gamma_{\tau}|\mathcal{F}_{\tau}^{(0)}\right)\right\| & \leq Ks \\ E\left(\sup_{u\in[0,s]}\left\|X'_{\tau+u}-X'_{\tau}\right\|^{q}|\mathcal{F}_{\tau}^{(0)}\right) & \leq Ks^{q/2} \\ E\left(\sup_{u\in[0,s]}\left\|c_{\tau+u}-c_{\tau}\right\|^{q}+\left\|\gamma_{\tau+u}-\gamma_{\tau}\right\|^{q}|\mathcal{F}_{\tau}^{(0)}\right) & \leq Ks^{(q/2)\wedge 1} \end{cases}$$

by Lemma 2.1.7, Corollary 2.1.9 in [33],

$$\text{(B.3)}\qquad \begin{cases} E\left(\sup_{u\in[0,s]}\left\|J_{\tau+u}-J_{\tau}\right\|^{q}|\mathcal{F}_{\tau}^{(0)}\right)\leq K_q\, sE\left[\widehat{\delta}(q)_{\tau,s}|\mathcal{F}_{\tau}^{(0)}\right] \\ E\left[\sup_{u\in[0,s\wedge 1]}\left(\frac{\|J_{\tau+u}-J_{\tau}\|}{s^{w}}\wedge 1\right)^{q}|\mathcal{F}_{\tau}^{(0)}\right]\leq K\, s^{(1-w\nu)(q/\nu\wedge 1)}a(s) \end{cases}$$

where $\widehat{\delta}(q)_{t,s}\equiv s^{-1}\int_t^{t+s}\int_E\|\delta(u,x)\|^q\,\lambda(\mathrm{d}x)\,\mathrm{d}u$ and $a(s)\to 0$ as $s\to 0$.
IV. Let $\varphi_n(t)=\sum_{h=1}^{l_n-1}\varphi_h^n\mathbb{1}_{((h-1)\Delta_n,h\Delta_n]}(t)$. For a generic process U, define

$$\text{(B.4)}\qquad U_{t,s}^n=\int_t^{t+s}\varphi_n(u-t)\,\mathrm{d}U_u$$

This quantity is useful in analyzing $\overline{U}_i^n$, with U being X or Y.
V. For $p\in\mathbb{N}^+$, $l,m=0,1$, by (4.1) in the main article and Riemann summation,

$$\text{(B.5)}\qquad \sum_{h=i}^{i+pl_n-2}\sum_{h'=h+1}^{i+pl_n-1}\phi_l\Big(\frac{h'-h}{l_n}\Big)\phi_m\Big(\frac{h'-h}{l_n}\Big)=l_n^2\left(p\Phi_{lm}-\Psi_{lm}\right)+O(pl_n)$$

VI. By Jensen's inequality and Doob's maximal inequality, we have the following lemma:

LEMMA 1. *Let $Z_i, i=1,\cdots,M$ be random variables, $\mathcal{H}_i=\sigma(Z_1,\cdots,Z_i)$ be the σ-algebra generated by $Z_1,\cdots,Z_i$, then*

$$\mathbb{E}\left(\sup_{m=1,\cdots,M}\left\|\sum_{i=1}^{m}\left[Z_i-E\left(Z_i|\mathcal{H}_i\right)\right]\right\|\right)\leq K\left(\sum_{i=1}^{M}\mathbb{E}\left(\|Z_i\|^2\right)\right)^{1/2}$$

APPENDIX C: DERIVATION FOR THEOREM 1

C.1. Properties of spot estimator: I. jumps. By Assumption A-ν, A-γ, (B.1), (B.2), (B.3), and $\overline{J}_i^n=\psi_n^{-1/2}\overline{J}^n_{i\Delta_n,(l_n-1)\Delta_n}$ from (B.4), we know that under the tunning (3.9)

$$\text{(C.1)}\qquad \begin{cases} E_i^n\left(\left\|\overline{Y}_i^{*n}\right\|^q\right) & \leq K_q\Delta_n^{q/2},\ \forall q\leq 8 \\ E_i^n\left(\left\|\overline{Y}_i^{n}\right\|^q\right) & \leq K_q\Delta_n^{(q/2)\wedge(q/4+1/2)},\ \forall q\leq 8 \\ E_i^n\left(\left\|\widehat{Y}_i^{*n}\right\|^q\vee\left\|\widehat{Y}_i^{n}\right\|^q\right) & \leq K_q\Delta_n^{q},\ \forall q\leq 4 \\ E_i^n\left[\left(\frac{\|\overline{J}_i^n\|}{\Delta_n^w}\wedge 1\right)^q\right] & \leq K_q\Delta_n^{[1/2-(w-1/4)\nu]\times[1\wedge(q/\nu)]}\,a_n,\ \forall q\leq 8 \end{cases}$$

for some $a_n\to 0$. We can write

$$\left\|\left(\overline{Y}_i^n\cdot\overline{Y}_i^{n,\mathrm{T}}\mathbb{1}_{\{\|\overline{Y}_i^n\|\leq\nu_n\}}-\widehat{Y}_i^n\right)-\left(\overline{Y}_i^{*n}\cdot\overline{Y}_i^{*n,\mathrm{T}}-\widehat{Y}_i^{*n}\right)\right\|\leq\sum_{r=1}^{3}\eta_i^{n,r}$$

where

$$\eta_i^{n,1} = \left\| \bar{Y}_i^n \cdot \bar{Y}_i^{n,\mathrm{T}} \mathbb{1}_{\{\|\bar{Y}_i^n\| \le \nu_n\}} - \bar{Y}_i^{*n} \cdot \bar{Y}_i^{*n,\mathrm{T}} \mathbb{1}_{\{\|\bar{Y}_i^{*n}\| \le \nu_n\}} \right\| \tag{C.2}$$

$$\eta_i^{n,2} = \left\| \widehat{Y}_i^n - \widehat{Y}_i^{*n} \right\|$$

$$\eta_i^{n,3} = \left\| \bar{Y}_i^{*n} \right\|^2 \mathbb{1}_{\{\|\bar{Y}_i^{*n}\| > \nu_n\}}$$

Let $u_n = \nu_n / \Delta_n^{1/2}$, $Z_i^n = \|\bar{Y}_i^{*n}\| / \Delta_n^{1/2}$, $Q_i^n = (\|\bar{J}_i^n\| / \Delta_n^{1/2}) \wedge 1$, $V_i^n = (\|\bar{J}_i^n\| / \Delta_n^{\rho}) \wedge 1$, we have

$$\Delta_n \eta_i^{n,1} \le u_n^{-2/(1-2\rho)} (Z_i^n)^{2+2/(1-2\rho)} + (1 + Z_i^n) \left[Q_i^n + u_n^2 (V_i^n)^2 \right]$$

By successive conditioning and (B.3), there is a sequence $a_n \to 0$ such that

$$E_i^n \left[(\eta_i^{n,1})^q \right] \le K_q \, \Delta_n^{2\rho q + 1/2 - (\rho - 1/4)\nu} a_n$$

Analyzing $\eta_i^{n,2}$ with (B.2), (B.3), analyzing $\eta_i^{n,3}$ with Cauchy-Schwarz inequality, Markov's inequality, (C.1), we can get the following lemma:

LEMMA 2. *Assume (B.1), (3.9), Assumption A-ν and A-γ, then $\exists\, a_n \to 0$ such that $\forall q \le 4$,*

$$E_i^n \left(\|\widehat{c}_i^n - \widehat{c}_i^{*n}\|^q \right) \le K_q \left(a_n \Delta_n^{1/2 - (\rho - 1/4)\nu - (1-2\rho)q} + \Delta_n^{1/2} \right)$$

C.2. Properties of spot estimator: II. continuous part.

C.2.1. *variables.* "If there is a rifle handing on the wall in act one, it must be fired in the next act. Otherwise it has no business being there", said the Russian playwright Anton Chekhov. Define

$$\bar{C}_i^n = \frac{1}{\psi_n} \sum_{h=1}^{l_n - 1} (\varphi_h^n)^2 \Delta_{i+h}^n C, \; C_t = \int_0^t c_s \, \mathrm{d}s$$

$$D_i^n = \bar{C}_i^n - c_i^n \Delta_n$$

$$\Gamma_h^n = \Gamma_{h,h}^n, \; \Gamma_{h,h'}^n = \frac{1}{\psi_n} \sum_{v = h \vee h'}^{h \wedge h' + l_n - 1} (\varphi_{v-h+1}^n - \varphi_{v-h}^n)(\varphi_{v-h'+1}^n - \varphi_{v-h'}^n) \gamma_v^n$$

$$R_i^n = \widehat{Y}_i^{*n} - \Gamma_i^n$$

$$\zeta_i^n = \bar{Y}_i^{*n} \cdot \bar{Y}_i^{*n,\mathrm{T}} - \bar{C}_i^n - \Gamma_i^n$$

given $p \in \mathbb{N}^+$, define

$$\zeta(W,p)_i^n = \sum_{h=i}^{i+pl_n-1} \left[(\sigma_i^n \bar{W}_h^n) \cdot (\sigma_i^n \bar{W}_h^n)^{\mathrm{T}} - \bar{C}_h^n \right]$$

$$\zeta(X,p)_i^n = \sum_{h=i}^{i+pl_n-1} \left(\bar{X}_h^n \cdot \bar{X}_h^{n,\mathrm{T}} - \bar{C}_h^n \right)$$

$$\zeta(X,p)_i'^n = \sum_{h=i}^{i+pl_n-2} \sum_{h'=h+1}^{i+pl_n-1} \bar{X}_h^n \cdot \bar{X}_{h'}^{n,\mathrm{T}} \phi_1\left(\frac{h'-h}{l_n} \right)$$

$$\zeta(p)_i^n = \sum_{h=i}^{i+pl_n-1} \zeta_h^n$$

Let $m(n,p) = \left\lfloor \frac{k_n - l_n}{(p+1)l_n} \right\rfloor$, $a(n,p,h) = 1 + h(p+1)l_n$, $b(n,p,h) = a(n,p,h) + pl_n$, then the estimation error of $\widehat{c}_i^{*n}$ can be decomposed as

$$\beta_i^n := \widehat{c}_i^{*n} - c_i^n = \xi_i^{n,0} + \xi_i^{n,1} + \xi_i^{n,2} + N(p)_i^n + M(p)_i^n \tag{C.3}$$

where

$$\xi_i^{n,0} = \frac{1}{k_n - l_n} \sum_{h=1}^{k_n - l_n + 1} c_{i+h}^n - c_i^n$$

$$\xi_i^{n,1} = \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n - l_n + 1} D_{i+h}^n$$

$$\xi_i^{n,2} = \frac{-1}{(k_n - l_n)\Delta_n} \sum_{h=1}^{k_n - l_n + 1} R_{i+h}^n$$

$$N(p)_i^n = \frac{1}{(k_n - l_n)\Delta_n} \Big(\sum_{h=0}^{m(n,p)-1} \zeta(1)_{i+b(n,p,h)}^n + \sum_{h=m(n,p)(p+1)l_n}^{k_n - l_n} \zeta_{i+1+h}^n \Big)$$

$$M(p)_i^n = \frac{1}{(k_n - l_n)\Delta_n} \sum_{h=0}^{m(n,p)-1} \zeta(p)_{i+a(n,p,h)}^n$$

C.2.2. *bounds on* $\|\xi_i^{n,r}\|$. By Assumption A-ν, (B.1), (B.2),

$$\begin{cases} \left\| E\left(\xi_i^{n,0} | \mathcal{F}_i^{(0),n}\right) \right\| \le K k_n \Delta_n \\ E\left(\|\xi_i^{n,0}\|^q | \mathcal{F}_i^{(0),n}\right) \le K_q (k_n \Delta_n)^{(q/2)\wedge 1},\ q \ge 0 \end{cases} \tag{C.4}$$

combined with (3.9),

$$\begin{cases} \left\| E\left(\xi_i^{n,1} | \mathcal{F}_i^{(0),n}\right) \right\| \le K \Delta_n^{1/2} \\ E\left(\|\xi_i^{n,1}\|^q | \mathcal{F}_i^{(0),n}\right) \le K_q \Delta_n^{[(q/2)\wedge 1]/2},\ q \in \mathbb{N}^+ \end{cases} \tag{C.5}$$

By Assumption A-γ,

$$\begin{aligned} &\left\| E_i^n\left(\xi_i^{n,2}\right) \right\| \le K \Delta_n^{-1} \\ &E_i^n\left(\|\xi_i^{n,2}\|^q\right) \le \begin{cases} K\, k_n^{-1/2},\ q = 1; \\ K_q \left(k_n^{-q+1} + k_n^{-q} \Delta_n^{-q/2+1} \right), q = 2,3,4 \end{cases} \end{aligned} \tag{C.6}$$

C.2.3. *estimates of* $\zeta(X,p)_i^n$ *&* $\zeta(X,p)_i'^n$. $\bar{C}_i^n = (\psi^n)^{-1} C^n_{i\Delta_n,(l_n-1)\Delta_n}$ in view of (B.4), hence by (B.1)

$$\|\bar{C}_i^n\| \le K \Delta_n \tag{C.7}$$

According to (3.1) in the main article we have $\bar{X}_i^n = -\psi_n^{-1/2} \sum_{h=0}^{l_n-1} (\varphi_{h+1}^n - \varphi_h^n)(X_{i+h}^n - X_i^n)$, then by (B.2)

$$E\left(\|\bar{X}_i^n\|^q | \mathcal{F}_i^{(0),n}\right) \le K_q\, \Delta_n^{q/2} \tag{C.8}$$

Adopt the argument for (5.21) in [12] in the multivariate setting, we have

$$\text{(C.9)}\qquad \begin{aligned} E\big(\bar{W}_h^n \bar{W}_{h'}^{n,\mathrm{T}} | \mathcal{F}_i^{(0),n}\big) &= \tfrac{l_n \Delta_n}{\psi_n} \phi_0\big(\tfrac{|h'-h|}{l_n}\big)\mathbf{I} + O_p(l_n^{-1/2}\Delta_n) \\ E\big(\|\bar{W}_h^n\|^{2m} | \mathcal{F}_i^{(0),n}\big) &= \Delta_n^m (2m-1)!! + O_p(l_n^{-1}\Delta_n^m),\ m \in \mathbb{N}^+ \end{aligned}$$

Let $U_i^n(p) = \sum_{h=i}^{i+pl_n-1} (\sigma_i^n \bar{W}_h^n)(\sigma_i^n \bar{W}_h^n)^{\mathrm{T}}$, $S_i^n(p) = \sum_{h=i}^{i+pl_n-1} \bar{C}_h^n$, then

$$\text{(C.10)}\quad \zeta(W,p)_i^{n,jk}\zeta(W,p)_i^{n,lm} = U_i^n(p)^{jk}U_i^n(p)^{lm} + S_i^n(p)^{jk}S_i^n(p)^{lm} - U_i^n(p)^{jk}S_i^n(p)^{lm} - U_i^n(p)^{lm}S_i^n(p)^{jk}$$

By (B.5), (C.7), (C.9), (C.10), and through similar arguments in section 5.3 of [12] with a modification for multi-dimension, and exploit the connection between $\zeta(W,p)_i^n$ and $\zeta(X,p)_i^n$, in view of (4.2) in the main article, we have the following lemma:

LEMMA 3. *Assume Assumption A-ν, (B.1), l_n satisfies (3.9), then*

$$E\left(\|\zeta(X,p)_i^n\|^4 | \mathcal{F}_i^{(0),n}\right) \le Kp^4\Delta_n^2$$

$$\left\|E\left[\zeta(X,p)_i^n | \mathcal{F}_i^{(0),n}\right]\right\| \le Kp\Delta_n$$

$$E\left[\zeta(X,p)_i'^n | \mathcal{F}_i^{(0),n}\right] = \frac{\theta^2 l_n}{\psi_n}(p\Phi_{01} - \Psi_{01})\, c_i^n + p^2\, O_p(\Delta_n^{1/4})$$

$$E\left[\zeta(X,p)_i^{n,jk}\zeta(X,p)_i^{n,lm} | \mathcal{F}_i^{(0),n}\right] = \frac{2\theta^4}{\psi_n^2}(p\Phi_{00} - \Psi_{00})\, \Sigma(c_i^n)^{jk,lm} + p^2\, O_p(\Delta_n^{5/4})$$

C.2.4. *estimates of* $\zeta(p)_i^n$. For $i \le h, h' \le i + pl_n - 1$, by (B.5)

$$\text{(C.11)}\qquad \Gamma_{h,h'}^n = \frac{1}{\psi_n l_n}\phi_1\Big(\frac{|h'-h|}{l_n}\Big)\gamma_i + O_p(\Delta_n^{5/4})$$

$$\sum_{h=i}^{i+pl_n-2}\sum_{h'=h+1}^{i+pl_n-1} \Gamma_{h,h'}^{n,jk}\Gamma_{h,h'}^{n,lm} = \frac{1}{\psi_n^2}\,(p\Phi_{11} - \Psi_{11})\,\gamma_i^{n,jk}\gamma_i^{n,lm} + p^2\, O_p(\Delta_n^{5/4})$$

Let $\xi_{h_1\cdots h_q}^{n,j_1\cdots j_q} = \prod_{v=1}^q \big(\bar{Y}_{h_v}^{n,j_v} - \bar{X}_{h_v}^{n,j_v}\big)$. By Assumption A-$\gamma$,

(C.12)

$$E\left[\big(\bar{Y}_h^{n,j} - \bar{X}_h^{n,j}\big)^q\big(\bar{Y}_{h'}^{n,k} - \bar{X}_{h'}^{n,k}\big)^r | \widetilde{\mathcal{F}}_{h\wedge h'-1}^n\right] = \begin{cases} 0 & q+r=1 \\ \Gamma_{h,h'}^{n,jk} & q=r=1 \\ O_p\big(l_n^{-7/2}\big)\mathbb{1}_{\{|h-h'|\le l_n\}} & q+r=3 \\ O_p(l_n^{-8}) & q+r=8 \end{cases}$$

$$E\left[\xi_{hhh'h'}^{n,jklm} | \widetilde{\mathcal{F}}_{h\wedge h'}^n\right] = \Gamma_h^{n,jk}\Gamma_{h'}^{n,lm} + \Gamma_{h,h'}^{n,jl}\Gamma_{h,h'}^{n,km} + \Gamma_{h,h'}^{n,jm}\Gamma_{h,h'}^{n,kl} + O_p\left(l_n^{-5}\right)$$

then by (C.8), (C.12)

$$E(\zeta_h^n | \widetilde{\mathcal{F}}_h^n) = \bar{X}_h^n \bar{X}_h^{n\,\mathrm{T}} - \bar{C}_h^n,\ \ \|E_h^n(\zeta_h^n)\| \le K\Delta_n^{3/2},\ \ E_h^n\left(\|\zeta_h^n\|^4\right) \le K\Delta_n^4$$

$$E\big(\zeta_h^{n,jk}\zeta_{h'}^{n,lm} | \widetilde{\mathcal{F}}_{h\wedge h'}^n\big) = \sum_{r=1}^3 \vartheta(r)_{h,h'}^{n,jk,lm} + \big(\bar{X}_h^{n,j}\bar{X}_h^{n,k} - \bar{C}_h^{n,jk}\big)\big(\bar{X}_{h'}^{n,l}\bar{X}_{h'}^{n,m} - \bar{C}_{h'}^{n,lm}\big)$$

where

$$\vartheta(1)^{n,jk,lm}_{h,h'} = \bar{X}^{n,j}_h \bar{X}^{n,l}_{h'} \xi^{n,km}_{hh'} + \bar{X}^{n,j}_h \bar{X}^{n,m}_{h'} \xi^{n,kl}_{hh'} + \bar{X}^{n,k}_h \bar{X}^{n,l}_{h'} \xi^{n,jm}_{hh'} + \bar{X}^{n,k}_h \bar{X}^{n,m}_{h'} \xi^{n,jl}_{hh'}$$

$$\vartheta(2)^{n,jk,lm}_{h,h'} = \xi^{n,jklm}_{hhh'h'} - \xi^{n,jk}_{hh} \Gamma^{n,lm}_{h'} - \xi^{n,lm}_{h'h'} \Gamma^{n,jk}_{h} + \Gamma^{n,jk}_{h} \Gamma^{n,lm}_{h'}$$

$$\vartheta(3)^{n,jk,lm}_{h,h'} = \bar{X}^{n,j}_h \xi^{n,klm}_{hh'h'} + \bar{X}^{n,k}_h \xi^{n,jlm}_{hh'h'} + \bar{X}^{n,l}_{h'} \xi^{n,jkm}_{hhh'} + \bar{X}^{n,m}_{h'} \xi^{n,jkl}_{hhh'}$$

Let $\Upsilon^{n,jk,lm}_{h,h'} = \Theta(\bar{X}^n_h \bar{X}^{n,\mathrm{T}}_{h'}, \Gamma^n_{h,h'})^{jk,lm}$ in light of (4.2) in the main article, then

$$E\left(\zeta^{n,jk}_h \zeta^{n,lm}_{h'} | \widetilde{\mathcal{F}}^n_{h\wedge h'}\right) = \left(\bar{X}^{n,j}_h \bar{X}^{n,k}_h - \bar{C}^{n,jk}_h\right)\left(\bar{X}^{n,l}_{h'} \bar{X}^{n,m}_{h'} - \bar{C}^{n,lm}_{h'}\right) + \Upsilon^{n,jk,lm}_{h,h'} + \Sigma(\Gamma^n_{h,h'})^{jk,lm} + O_p\left(l_n^{-5} + (\|\bar{X}^n_h\| + \|\bar{X}^n_{h'}\|) l_n^{-7/2}\right)$$

hence

$$\zeta(p)^{n,jk}_i \zeta(p)^{n,lm}_i = \zeta(X,p)^{n,jk}_i \zeta(X,p)^{n,lm}_i + \sum_{h=i}^{i+pl_n-2} \sum_{h'=h+1}^{i+pl_n-1} \left[\Upsilon^{n,jk,lm}_{h,h'} + \Upsilon^{n,lm,jk}_{h,h'} + \Sigma(\Gamma^n_{h,h'})^{jk,lm} + \Sigma(\Gamma^n_{h,h'})^{lm,jk}\right] + \sum_{h=i}^{i+pl_n-1} \left[\Upsilon^{n,jk,lm}_{h,h'} + \Sigma(\Gamma^n_{h,h'})^{jk,lm}\right] + p^2\, O_p(\Delta_n^{5/4})$$

then by (C.8), (C.11)

$$E\left[\zeta(p)^{n,jk}_i \zeta(p)^{n,lm}_i | \widetilde{\mathcal{F}}^n_i\right] = \zeta(X,p)^{n,jk}_i \zeta(X,p)^{n,lm}_i + \frac{2}{\psi_n l_n} \Theta(\zeta(X,p)'^n_i, \gamma^n_i)^{jk,lm} + \frac{2}{\psi_n^2}(p\Phi_{11} - \Psi_{11})\, \Sigma(\gamma^n_i)^{jk,lm} + p^2\, O_p(\Delta_n^{5/4})$$

According to these results and Lemma 3, one can get the following lemma:

LEMMA 4. *Assume Assumption A-ν, A-γ, (B.1), l_n satisfies (3.9), then*

$$E[\zeta(p)^n_i | \widetilde{\mathcal{F}}^n_i] = \zeta(X,p)^n_i$$

$$\|E^n_i[\zeta(p)^n_i]\| \le Kp\Delta_n$$

$$E^n_i(\|\zeta(p)^n_i\|^q) \le K_q\, p^{\lfloor q/2 \rfloor \vee 1} \Delta_n^{q/2}, \quad q = 1,2,3,4$$

moreover

$$\left| E^n_i\left[\zeta(p)^{n,jk}_i \zeta(p)^{n,lm}_i\right] - (p+1)\theta\Delta_n\, \Xi(c^n_i, \gamma^n_i; p)^{jk,lm} \right| \le K\Delta_n^{5/4}$$

where

$$\Xi(x,z;p) = \frac{p\Phi_{00} - \Psi_{00}}{(p+1)\Phi_{00}}\, \Sigma(x) + \frac{p\Phi_{01} - \Psi_{01}}{(p+1)\Phi_{01}}\, \Theta(x,z) + \frac{p\Phi_{11} - \Psi_{11}}{(p+1)\Phi_{11}}\, \Sigma(z) \tag{C.13}$$

Let $p \asymp \Delta_n^{-1/12}$, based on Lemma 4,

$$\|E^n_i\,[M(p)^n_i]\| \le K\Delta_n^{1/2}$$

$$\|E_i^n\left[N(p)_i^n\right]\| \leq Kp^{-1}\Delta_n^{1/2}$$

$$E_i^n\left[\|M(p)_i^n\|^q\right] \leq \begin{cases} K_q\left(k_n\Delta_n^{1/2}\right)^{-q/2}, & q=1,2,4 \\ K\left(k_n\Delta_n^{1/2}\right)^{-2}, & q=3 \end{cases} \tag{C.14}$$

$$E_i^n\left[\|N(p)_i^n\|^q\right] \leq \begin{cases} Kp\left(k_n\Delta_n^{1/2}\right)^{-1}, & q=1 \\ K_q p^{-q/2}\left(k_n\Delta_n^{1/2}\right)^{-q/2}, & q=2,4 \\ Kp^{-1}\left(k_n\Delta_n^{1/2}\right)^{-2}, & q=3 \end{cases}$$

C.2.5. *estimates of* β_i^n. We need to define more variables:

$$\begin{array}{ll} \zeta(p)_{i,h}^n = \zeta(p)_{i+a(n,p,h)}^n & A(p)_{i+v}^n = \sum_{h=v}^{v+pl_n-1}(c_{i+h}^n - c_i^n)\Delta_n \\ D(p)_{i+v}^n = \sum_{h=v}^{v+pl_n-1} D_{i+h}^n & R(p)_{i+v}^n = \sum_{h=v}^{v+pl_n-1} R_{i+h}^n \end{array} \tag{C.15}$$

we have

TABLE 2
Estimates of ingredients under (3.9)

scaling properties	$E(\|\cdot\|^2\|\mathcal{F}_i^n)$	$\|E(\cdot\|\mathcal{F}_i^n)\|$
$R(p)_{i+v}^n$	$p\Delta_n^{3/2}$	$p\Delta_n^{3/2}$
$D(p)_{i+v}^n$	$p\Delta_n^{3/2}$	$p\Delta_n$
$A(p)_{i+v}^n$	$p^2\Delta_n^2(p\Delta_n^{-1/2}+v)$	$p\Delta_n^{3/2}(p\Delta_n^{-1/2}+v)$
$\zeta(p)_{i,h}^n$	$p\Delta_n$	$p\Delta_n$

Define

$$\alpha(p)_{i,h}^n = -R(p+1)_{i+a(n,p,h)}^n + D(p+1)_{i+a(n,p,h)}^n + A(p+1)_{i+a(n,p,h)}^n + \zeta(p+1)_{i,h}^n$$

By Table 2 and Cauchy-Schwarz inequality,

$$E_i^n\left(\left|\alpha(p)_{i,h}^{n,jk}\alpha(p)_{i,h}^{n,lm} - \zeta(p+1)_{i,h}^{n,jk}\zeta(p+1)_{i,h}^{n,lm}\right|\right) \leq K\left(p^2\Delta_n^{5/4} + p^{3/2}\Delta_n^{3/2}v^{1/2}\right) \tag{C.16}$$

Given $j,k,l,m=1,\cdots,d$, in the light of (C.3) and (C.15), by Table 2 we have

$$\left|E_i^n\left(\beta_i^{n,jk}\beta_i^{n,lm}\right) - (k_n\Delta_n^{1/2})^{-1}\Xi(c_i^n,\gamma_i)^{jk,lm}\right| = \sum_{r=1}^5 \mu_i^{n,r} + p\,O_p((k_n^2\Delta_n)^{-1})$$

where

$$\mu_i^{n,1} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-1} E_i^n\left(\left|\alpha(p)_{i,h}^{n,jk}\alpha(p)_{i,h}^{n,lm} - \zeta(p+1)_{i,h}^{n,jk}\zeta(p+1)_{i,h}^{n,lm}\right|\right)$$

$$\mu_i^{n,2} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-2}\sum_{h'=h+1}^{m(n,p)-1}\left|E_i^n\left[\alpha(p)_{i,h}^{n,jk}\alpha(p)_{i,h'}^{n,lm} + \alpha(p)_{i,h}^{n,lm}\alpha(p)_{i,h'}^{n,jk}\right]\right|$$

$$\mu_i^{n,3} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-1} E_i^n\left(\left|\zeta(p+1)_{i,h}^{n,jk}\zeta(p+1)_{i,h}^{n,lm}\right.\right.$$

$$-(p+2)\theta\Delta_n\,\Xi\big(c^n_{i+a(n,p,h)},\gamma^n_{i+a(n,p,h)};p+1\big)^{jk,lm}\Big|\Big)$$

$$\mu_i^{n,4}=\frac{(p+2)\theta}{(k_n-l_n)^2\Delta_n}\sum_{h=0}^{m(n,p)-1}\Big|E_i^n\Big[\Xi\big(c^n_{i+a(n,p,h)},\gamma^n_{i+a(n,p,h)};p+1\big)^{jk,lm}$$
$$-\Xi(c_i^n,\gamma_i^n;p+1)^{jk,lm}\Big]\Big|$$

$$\mu_i^{n,5}=\frac{1}{(k_n-l_n)\Delta_n^{1/2}}\left|\frac{(p+2)\theta}{(k_n-l_n)\Delta_n^{1/2}}\left\lfloor\frac{k_n}{(p+1)l_n}\right\rfloor-\frac{(k_n-l_n)}{k_n}\right|\cdot\Big|\Xi(c_i^n,\gamma_i^n;p+1)^{jk,lm}\Big|$$
$$+\frac{1}{k_n\Delta_n^{1/2}}\Big|\Xi(c_i^n,\gamma_i^n;p+1)^{jk,lm}-\Xi(c_i^n,\gamma_i^n)^{jk,lm}\Big|$$

Use Table 2 and (C.16) to get bounds on $\mu_i^{n,r}$, $r=1,2,3,4,5$; combine (C.3), (C.4), (C.5), (C.6), (C.14), we get the following lemma:

LEMMA 5. *Assume (B.1), (3.9) and Assumption A-ν, A-γ, given $p\in\mathbb{N}^+$,*

$$\|E_i^n(\beta_i^n)\|\le Kk_n\Delta_n$$
$$E_i^n(\|\beta_i^n\|^q)\le\begin{cases}K_q\Big[(k_n\Delta_n)^{(q/2)\wedge1}+(k_n\Delta_n^{1/2})^{-q/2}\Big], & q=1,2,4\\ Kk_n\Delta_n, & q=3\end{cases}$$

additionally

$$\Big|E_i^n\big(\beta_i^{n,jk}\beta_i^{n,lm}\big)-(k_n\Delta_n)^{-1/2}\Xi(c_i^n,\gamma_i^n)^{jk,lm}\Big|\le K\big[k_n\Delta_n+p^{-1}(k_n\Delta_n^{1/2})^{-1}\big]$$

C.3. Structure. Define

$$\lambda(x,z)=\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(x)\times\Xi(x,z)^{jk,lm}$$
$$\eta_i^n=\lambda(\widehat{c}_i^n,\widehat{\gamma}_i^n)-\lambda(\widehat{c}_i^{*n},\widehat{\gamma}_i^n)$$

As $\Delta_n^{-1/4}|a_t^n-1|<k_n\Delta_n^{3/4}\to0$, letting $a_t^n=1$ doesn't not affect the asymptotic analysis. By Cramér-Wold theorem, we can suppose g is $\mathbb{R}$-valued. We have

$$\Delta_n^{-1/4}\big[\widehat{S}(g)^n-S(g)\big]=\overline{S}^{n,0}+\overline{S}^{n,1}+\overline{S}(p)^{n,2}+\overline{S}^{n,3}+\overline{S}(p)^{n,4}\tag{C.17}$$

where

$$\overline{S}_t^{n,0}=\Delta_n^{-1/4}\left[\sum_{i=0}^{N_t^n-1}\int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n}g(c^n_{ik_n})-g(c_s)\,\mathrm{d}s-\int_{N_t^nk_n\Delta_n}^{t}g(c_s)\,\mathrm{d}s\right]$$

$$\overline{S}_t^{n,1}=k_n\Delta_n^{3/4}\sum_{i=0}^{N_t^n-1}\Big[g(\widehat{c}^n_{ik_n})-g(\widehat{c}^{*n}_{ik_n})-(2k_n\Delta_n^{1/2})^{-1}\eta^n_{ik_n}\Big]$$

$$\overline{S}(p)_t^{n,2}=k_n\Delta_n^{3/4}\sum_{i=0}^{N_t^n-1}\sum_{j,k=1}^{d}\partial_{jk}g(c^n_{ik_n})\left[\sum_{r=0}^{2}\xi^{n,r,jk}_{ik_n}+N(p)^{n,jk}_{ik_n}\right]$$

$$\overline{S}_t^{n,3} = k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n-1} \Big[g(\widehat{c}_{ik_n}^{*n}) - g(c_{ik_n}^n) - \sum_{j,k=1}^{d} \partial_{jk} g(c_{ik_n}^n) \beta_{ik_n}^{n,jk} - (2k_n \Delta_n^{1/2})^{-1} \lambda(\widehat{c}_i^{*n}, \widehat{\gamma}_i^n) \Big]$$

$$\overline{S}(p)_t^{n,4} = k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n-1} \sum_{j,k=1}^{d} \partial_{jk} g(c_{ik_n}^n) \times M(p)_{ik_n}^{n,jk}$$

C.4. Asymptotic negligibility. First of all, we need to get bounds on $\partial^r g(\widehat{c}_i^n)$, $r \le 3$, where $\partial^r g$ denotes the r-th order partial derivatives. Let $\overline{c}_i^n = (k_n - l_n)^{-1} \sum_{h=1}^{k_n - l_n + 1} c_i^n$ and $I_t^n = \{0, 1, \cdots, N_t^n - 1\}$, note that $|I_t^n| \asymp (k_n \Delta_n)^{-1}$, according to Lemma 2, there is a sequence $a_n \to 0$ such that

$$\mathbb{E}\Big(\sup_{i \in I_n} \|\widehat{c}_i^n - \widehat{c}_i^{*n}\| \Big) \le K \left(a_n \Delta_n^{\kappa - 1/2 - (\rho - 1/4)\nu - (1-2\rho)} + \Delta_n^{\kappa - 1/2} \right)$$

Note $\widehat{c}_i^{*n} - \overline{c}_i^n = \xi_i^{n,1} + \xi_i^{n,2} + N(p)_i^n + M(p)_i^n$, by (C.5), (C.6), (C.14) and $\kappa < 3/4$, $E_i^n \big(\|\widehat{c}_i^{*n} - \overline{c}_i^n\|^4 \big) \le K \Delta_n^{2\kappa - 1}$, so

$$\mathbb{E}\Big(\sup_{i \in I_n} \|\widehat{c}_i^{*n} - \overline{c}_i^n\|^4 \Big) \le K \Delta_n^{3\kappa - 2}$$

hence by (3.9) and Markov's inequality

$$\sup_{i \in I_n} \|\widehat{c}_i^n - \overline{c}_i^n\| = o_p(1) \tag{C.18}$$

According to (B.1) and convexity, $\overline{c}_i^n \in \mathcal{S}$. By (C.18), $\widehat{c}_i^n \in \mathcal{S}^\epsilon$ when n is sufficiently large. Therefore by (3.8) in the main article, for asymptotic analysis we can assume

$$\|\partial^r g(\widehat{c}_i^n)\| \le K, \quad \forall r = 0, 1, 2, 3, \forall i \in I_t^n \tag{C.19}$$

Through an almost identical argument for Lemma 4.4 in [21],

$$\overline{S}^{n,0} \xrightarrow{u.c.p.} 0 \tag{C.20}$$

Define function g_n on $\mathcal{S}_d^+ \times \mathcal{S}_d^+$ as $g_n(x, z) = g(x) - (2k_n \Delta_n^{1/2})^{-1} \lambda(x, z)$. Based on (C.19),

$$\|g_n(x,z) - g_n(y,z)\| \le K\|x - y\| + K(k_n \Delta_n^{1/2})^{-1} \|x - y\| \left(\|x\|^2 + \|z\|^2 + \|x - y\|^2 + \|z\| \|x - y\| \right)$$

so $\|g_n(x,z) - g_n(y,z)\| \le K\|x - y\|$ when n is sufficiently large. By Lemma 2

$$\mathbb{E}\left(\sup_{u \in [0,t]} \left\| \overline{S}_u^{n,1} \right\| \right) \le k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n - 1} \left\| g_n(\widehat{c}_{ik_n}^n, \widehat{\gamma}_{ik_n}^n) - g_n(\widehat{c}_{ik_n}^{*n}, \widehat{\gamma}_{ik_n}^n) \right\| \le Kt \left(a_n \Delta_n^{1/4 - (\rho - 1/4)\nu - (1 - 2\rho)} + \Delta_n^{1/4} \right)$$

Since $\rho > \frac{3-\nu}{4(2-\nu)}$, $1/4 - (\rho - 1/4)\nu - (1 - 2\rho) > 0$, we have the following lemma:

LEMMA 6. *Assume Assumption A-ν, A-γ, (3.9), (C.19) then*

$$\overline{S}^{n,1} \xrightarrow{u.c.p.} 0$$

Given $e_i^n \in \mathbb{R}^{d\times d}$, consider the process

$$\overline{S}_t^{*n} = k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n-1} \sum_{j,k=1}^{d} \partial_{jk} g(c_{ik_n}^n) \times e_{ik_n}^{n,jk}$$

suppose e_i^n satisfies

$$\begin{cases} \|E_i^n(e_i^n)\| & \le K\Delta_n^{1/4} a_n \\ E_i^n\left(\|e_i^n\|^2\right) & \le K(k_n\Delta_n^{1/2})^{-1} b_n \end{cases} \tag{C.21}$$

where $a_n, b_n \to 0$. Since ∂g is bounded by (B.1),

$$\mathbb{E}\left(\sup_{u\in[0,t]} \left\|\overline{S}_u^{*n}\right\|\right) \le K k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n-1} \mathbb{E}\left(\left\|E_{ik_n}^n\left(e_{ik_n}^n\right)\right\|\right)$$
$$+ K k_n \Delta_n^{3/4} \mathbb{E}\left(\sup_{s\in[0,t]} \left\| \sum_{i=0}^{N_s^n-1} \left[e_{ik_n}^n - E_{ik_n}^n\left(e_{ik_n}^n\right)\right] \right\|\right)$$

by Lemma 1,

$$\mathbb{E}\left(\sup_{s\in[0,t]} \left\| \sum_{i=0}^{N_s^n-1} \left[e_{ik_n}^n - E_{ik_n}^n\left(e_{ik_n}^n\right)\right] \right\|\right) \le K \left(\sum_{i=0}^{N_t^n-1} \mathbb{E}\left(\|e_{ik_n}^n\|^2\right)\right)^{1/2}$$

note that $k_n\Delta_n N_t^n \asymp t$, we have

$$\mathbb{E}\left(\sup_{u\in[0,t]} \left\|\overline{S}_u^{*n}\right\|\right) \le K\left(ta_n + \sqrt{tb_n}\right) \to 0$$

To show the asymptotic negligibility of $\overline{S}(p)^{n,2}$, we need to show e_i satisfies (C.21) in each of the following 4 cases:

① when $e_i^n = \xi_i^{n,0}$, by (C.4), $a_n = k_n\Delta_n^{3/4}$, $b_n = (k_n\Delta_n^{3/4})^2$;
② when $e_i^n = \xi_i^{n,1}$, by (C.5), $a_n = \Delta_n^{1/4}$, $b_n = k_n\Delta_n$;
③ when $e_i^n = \xi_i^{n,2}$, by (C.6), $a_n = \Delta_n^{3/4}$, $b_n = \Delta_n^{1/2}$;
④ when $e_i^n = N(p)_i^n$, by (C.14), $a_n = p^{-1}\Delta_n^{1/4}$, $b_n = p^{-1}$.

Hence we have the following lemma:

LEMMA 7. *Assume Assumption A-ν, A-γ, (3.9), (C.19), and let $p \asymp \Delta_n^{-1/12}$, then*

$$\overline{S}(p)^{n,2} \xrightarrow{u.c.p.} 0$$

Let $\chi_i^n = \widehat{\gamma}_i^n - \gamma_i^n$. By (B.2), the choice of m_n and Jensen's inequality

$$E_i^n(\|\chi_i^n\|^q) \le K_q \Delta_n^{q/4},\ q = 1, 2 \tag{C.22}$$

Let $\iota_i^n = \lambda(\widehat{c}_i^{*n}, \widehat{\gamma}_i^n) - \lambda(c_i^n, \gamma_i^n)$, then by (4.2) and (4.3) in the main article

$$\|\iota_i^n\| \le K\left(\|\beta_i^n\| + \|\chi_i^n\| + \|\beta_i^n\|^2 + \|\beta_i^n\|\|\chi_i^n\| + \|\chi_i^n\|^2\right)$$

hence by Lemma 5, (3.9), (C.22),

$$E_i^n(\|\iota_i^n\|) \le K k_n^{-1/2} \Delta_n^{-1/4} \tag{C.23}$$

We can rewrite $\overline{S}^{n,3}$ as

$$\overline{S}^{n,3} = G^n + H^n$$

where

$$G_t^n = k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n - 1} \left[s_{ik_n}^n + \frac{1}{2k_n \Delta_n^{1/2}} \iota_{ik_n}^n + E_{ik_n}^n \left(v_{ik_n}^n \right) \right]$$

$$H_t^n = k_n \Delta_n^{3/4} \sum_{i=0}^{N_t^n - 1} \left[v_{ik_n}^n - E_{ik_n}^n \left(v_{ik_n}^n \right) \right]$$

$$s_i^n = g(c_i^n + \beta_i^n) - g(c_i^n) - \sum_{j,k=1}^{d} \partial_{jk} g(c_i^n) \beta_i^{n,jk} - \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(c_i^n) \beta_i^{n,jk} \beta_i^{n,lm}$$

$$v_i^n = \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(c_i^n) \left[\beta_i^{n,jk} \beta_i^{n,lm} - (k_n \Delta_n^{1/2})^{-1} \Xi(c_i^n, \gamma_i^n)^{jk,lm} \right]$$

By Lemma 5, (C.19), (C.23), if we let $p \asymp \Delta_n^{-12}$,

$$\mathbb{E}\left(\sup_{s \in [0,t]} \|G_s^n\| \right) \leq Kt \left[k_n \Delta_n^{3/4} + (k_n \Delta_n^{2/3})^{-1} \right] \tag{C.24}$$

and

$$E_i^n(\|v_i^n\|^2) \leq K \sum_{j,k,l,m=1}^{d} E_i^n \left(\left| \beta_i^{n,jk} \beta_i^{n,lm} - \frac{1}{k_n \Delta_n^{1/2}} \Xi(c_i^n, \gamma_i^n)^{jk,lm} \right|^2 \right) \leq K \left[k_n \Delta_n + (k_n \Delta_n^{1/2})^{-2} \right]$$

then Lemma 1 implies

$$\mathbb{E}\left(\sup_{s \in [0,t]} \|H_s^n\| \right) \leq K k_n \Delta_n^{3/4} \left(\sum_{i=0}^{N_t^n - 1} \mathbb{E}(\|v_{ik_n}^n\|^2) \right)^{1/2} \leq K\sqrt{t} \left[k_n \Delta_n^{3/4} + (k_n \Delta_n^{1/2})^{-1/2} \right] \tag{C.25}$$

According to (C.24) and (C.25), we have the following lemma:

LEMMA 8. *Assume Assumption A-ν, A-γ, (3.9), (C.19), then*

$$\overline{S}^{n,3} \xrightarrow{u.c.p.} 0$$

C.5. Stable convergence in law to a continuous Itô semimartingale. Recall (C.15), let

$$\Lambda(p)_{i,h}^n = \partial g(c_{ik_n}^n) \odot \zeta(p)_{ik_n,h}^n \tag{C.26}$$

where $\odot$ denotes Hadamard product, i.e., $\Lambda(p)_{i,h}^{n,jk} = \partial_{jk} g(c_{ik_n}^n) \odot \zeta(p)_{ik_n,h}^{n,jk}$. We can write

$$\overline{S}(p)_t^{n,4} = \frac{k_n}{k_n - l_n} \sum_{j,k=1}^{d} \Delta_n^{-1/4} \sum_{i=0}^{N_t^n - 1} \sum_{h=0}^{m(n,p)-1} \Lambda(p)_{i,h}^{n,jk}$$

Let $\mathcal{H}(p)^n_{i,h} = \mathcal{F}^n_{ik_n+a(n,p,h)}$, by Lemma 4,

$$\Delta_n^{-1/2} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \left\| E\left[\zeta(p)^n_{ik_n,h} | \mathcal{H}(p)^n_{i,h}\right] \right\|^2 \leq Ktp\Delta_n$$

Let N be a bounded martingale orthogonal to W or $N = W^l$ for some $l = 1, \cdots, d'$, and $\Delta N(p)^n_{i,h} = N^n_{ik_n+b(n,p,h)} - N^n_{ik_n+a(n,p,h)}$. The following four statements about convergence in probability for any indices j, k, l, m can verify the conditions of theorem IX.7.28 in [34]:

$$\Delta_n^{-1/4} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \left\| E\left[\Lambda(p)^n_{i,h} | \mathcal{H}(p)^n_{i,h}\right] \right\| \xrightarrow{\mathbb{P}} 0 \tag{C.27}$$

$$\Delta_n^{-1} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} E\left[\left\| \Lambda(p)^n_{i,h} \right\|^4 | \mathcal{H}(p)^n_{i,h} \right] \xrightarrow{\mathbb{P}} 0 \tag{C.28}$$

$$\Delta_n^{-1/4} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \left\| E\left[\Lambda(p)^n_{i,h} \Delta N(p)^n_{i,h} | \mathcal{H}(p)^n_{i,h}\right] \right\| \xrightarrow{\mathbb{P}} 0 \tag{C.29}$$

$$\Delta_n^{-1/2} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} E\left[\Lambda(p)^{n,jk}_{i,h} \Lambda(p)^{n,lm}_{i,h} | \mathcal{H}(p)^n_{i,h} \right] \xrightarrow{\mathbb{P}} \int_0^t \partial_{jk} g(c_s) \partial_{lm} g(c_s)^{\mathrm{T}} \Xi(c_s, \gamma_s; p)^{jk,lm} \, \mathrm{d}s \tag{C.30}$$

Under (C.19), one can verify (C.27), (C.28) by the second and third claims of Lemma 4, respectively. The same argument as that for (5.58) in [12] leads to (C.29). By the last claim of Lemma 4, the left-hand side of (C.30) equals

$$\sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \partial_{jk} g(c^n_{ik_n}) \partial_{lm} g(c^n_{ik_n})^{\mathrm{T}} \times \Xi\big(c^n_{ik_n+a(n,p,h)}, \gamma^n_{ik_n+a(n,p,h)}; p\big)^{jk,lm} (p+1) l_n \Delta_n + tp\, O_p(\Delta_n^{1/4})$$

then (C.30) is verified by Riemann summation. By Theorem IX.7.28 in [34] we have the following lemma:

LEMMA 9. *Assume Assumption A-ν, A-γ, (3.9), (C.19), then for $\forall p \in \mathbb{N}^+$,*

$$\overline{S}(p)^{n,4} \xrightarrow{\mathcal{L}-s(f)} Z(p)$$

where $Z(p)$ is a process defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, such that conditioning on $\mathcal{F}$ it is a mean-0 continuous Itô martingale with variance

$$\widetilde{E}[Z(p)Z(p)^{\mathrm{T}} | \mathcal{F}] = \int_0^t \sum_{j,k,l,m=1}^{d} \partial_{jk} g(c_s) \partial_{lm} g(c_s)^{\mathrm{T}} \Xi(c_s, \gamma_s; p)^{jk,lm} \, \mathrm{d}s$$

where $\widetilde{E}$ is the conditional expectation operator on the extended probability space and $\Xi(x, z; p)$ is defined in (C.13).

By (C.20), Lemma 6, 7, 8, 9, and $\Xi(x, z; p) \to \Xi(x, z)$ as $p \to \infty$, we arrive at the asymptotic result in Theorem 1.

APPENDIX D: DERIVATION FOR THEOREM 2

As a counterpart to (C.1), we have under (3.12)

$$\text{(D.1)}\qquad \begin{cases} E_i^n\left(\left\|\bar{Y}_i^{*n}\right\|^q\right) & \leq K_q\Delta_n^{q/2},\ \forall q\leq 8\\ E_i^n\left(\left\|\bar{Y}_i^{n}\right\|^q\right) & \leq K_q\Delta_n^{(q/2)\wedge[1/2-\delta+(1/2+\delta)q/2]},\ \forall q\leq 8\\ E_i^n\left[\left(\frac{\|\bar{J}_i^n\|}{\Delta_n^w}\wedge 1\right)^q\right] & \leq K_q\Delta_n^{[1/2-\delta-(w-1/4-\delta/2)\nu]\times[1\wedge(q/\nu)]}a_n,\ \forall q\leq 8\end{cases}$$

for some $a_n\to 0$. It follows that

$$\left\|\bar{Y}_i^n\cdot\bar{Y}_i^{n,\mathrm{T}}\mathbb{1}_{\{\|\bar{Y}_i^n\|\leq\nu_n\}}-\bar{Y}_i^{*n}\cdot\bar{Y}_i^{*n,\mathrm{T}}\right\|\leq\eta_i^{n,1}+\eta_i^{n,3}$$

where $\eta_i^{n,1}$ and $\eta_i^{n,3}$ are defined in (C.2).
Through a similar argument to that in Section C.1, for $1\leq q\leq 4$ there is a sequence $a_n\to 0$ such that under (3.12),

$$E_i^n\left[(\eta_i^{n,1})^q\right]\leq K_q\left[\Delta_n^{2q}+a_n\Delta_n^{2q\rho+1/2-\delta-(\rho-1/4-\delta/2)\nu}\right]$$
$$E_i^n\left[(\eta_i^{n,3})^q\right]\leq K_{q,q'}\Delta_n^{q+q'(1-2\rho)/2},\ q'>1/(1-2\rho)$$

Note $\|\widetilde{c}_i^n-\widetilde{c}_i^{*n}\|\leq[(k_n-l_n)\Delta_n]^{-1}\sum_{h=1}^{k_n-l_n+1}\left[\eta_{i+h}^{n,1}+\eta_{i+h}^{n,3}\right]$, then

$$\text{(D.2)}\qquad E_i^n\left(\|\widetilde{c}_i^n-\widetilde{c}_i^{*n}\|^q\right)\leq K_q\left(a_n\Delta_n^{1/2-\delta-(\rho-1/4-\delta/2)\nu-(1-2\rho)q}+\Delta_n^q\right)$$

Define the following quantities

$$\xi_i'^{n,2}=\frac{1}{(k_n-l_n)\Delta_n}\sum_{h=1}^{k_n-l_n+1}\Gamma_{i+h}^n$$
$$\beta_i'^n=\xi_i^{n,0}+\xi_i^{n,1}+\xi_i'^{n,2}+N(p)_i^n+M(p)_i^n$$

where $\xi_i^{n,0}$, $\xi_i^{n,1}$, $N(p)_i^n$, $M(p)_i^n$ are defined in (C.3). Note $\widetilde{c}_i^{*n}-c_i^n=\beta_i'^n$.
Under the tuning (3.12),

$$\text{(D.3)}\qquad \begin{cases}\left\|E\left(\xi_i^{n,1}|\mathcal{F}_i^{(0),n}\right)\right\| & \leq K\Delta_n^{1/2-\delta}\\ E\left(\|\xi_i^{n,1}\|^q|\mathcal{F}_i^{(0),n}\right) & \leq K_q\,\Delta_n^{[(q/2)\wedge 1](1/2-\delta)},\ q\in\mathbb{N}^+\end{cases}$$

By (3.1) and Assumption A-γ in the main article,

$$\text{(D.4)}\qquad \begin{aligned}\left\|E_i^n\left(\xi_i'^{n,2}\right)\right\| &\leq K\Delta_n^{2\delta}\\ E_i^n\left(\|\xi_i'^{n,2}\|^q\right) &\leq K_q\,\Delta_n^{2\delta q}\end{aligned}$$

Based on the same derivations in Section C.2 that lead to Lemma 4, we know

$$\|E_i^n\left[N(p)_i^n\right]\|\leq K\left(p^{-1}\Delta_n^{1/2-\delta}+pk_n^{-1}\Delta_n^{-2\delta}\right)$$

$$\text{(D.5)}\qquad E_i^n\left[\|N(p)_i^n\|^q\right]\leq\begin{cases}K_q\,p^{-q/2}\left(k_n\Delta_n^{1/2+\delta}\right)^{-q/2}, & q=1,2,4\\ Kp^{-1}\left(k_n\Delta_n^{1/2+\delta}\right)^{-2}, & q=3\end{cases}$$

and

$$\|E_i^n\left[M(p)_i^n\right]\|\leq K\Delta_n^{1/2-\delta}$$

$$\text{(D.6)}\qquad E_i^n\left[\|M(p)_i^n\|^q\right]\leq\begin{cases}K_q\left(k_n\Delta_n^{1/2+\delta}\right)^{-q/2}, & q=1,2,4\\ K\left(k_n\Delta_n^{1/2+\delta}\right)^{-2}, & q=3\end{cases}$$

Since $\widetilde{c}_i^{*n} - \overline{c}_i^n = \xi_i^{n,1} + \xi_i'^{n,2} + N(p)_i^n + M(p)_i^n$, based on (D.3), (D.4), (D.5), (D.6),

$$\mathbb{E}\Big(\sup_{i\in I_n} \|\widetilde{c}_i^{*n} - \overline{c}_i^n\|^4\Big) \le K\Delta_n^{3\kappa-2-2\delta}$$

then use $\|\widetilde{c}_i^n - \overline{c}_i^n\| \le \|\widetilde{c}_i^n - \widetilde{c}_i^{*n}\| + \|\widetilde{c}_i^{*n} - \overline{c}_i^n\|$ and (D.2) with $q=1$, under (3.12)

$$\sup_{i\in I_n} \|\widetilde{c}_i^n - \overline{c}_i^n\| = o_p(1)$$

Therefore, just as explained at the beginning of Section C.4, when the p.s.d. plug-in $\widetilde{c}_i^n$ and the tuning (3.12) are used, we can still assume (C.19).
We know by combining (C.4), (D.3), (D.4), (D.5) and (D.6),

$$\big\|E_i^n\left(\beta_i'^n\right)\big\| \le K\big(k_n\Delta_n + \Delta_n^{2\delta}\big)$$

$$\text{(D.7)}\quad E_i^n\left(\|\beta_i'^n\|^q\right) \le \begin{cases} K_q\big[(k_n\Delta_n)^{(q/2)\wedge 1} + \Delta_n^{[(q/2)\wedge 1](1/2-\delta)} \\ \qquad\qquad + \Delta^{2\delta q} + \big(k_n\Delta_n^{1/2+\delta}\big)^{-q/2}\big], & q=1,2,4 \\ K\big[k_n\Delta_n + \Delta_n^{1/2-\delta} + \big(k_n\Delta_n^{1/2+\delta}\big)^{-2}\big], & q=3 \end{cases}$$

Now we study the quadratic form of $\beta_i'^n$. Recall the definitions in (C.15) and the estimates in Table 2. Let $R(p)_{i+v}'^n = \sum_{h=v}^{v+pl_n-1}\Gamma_{i+h}^n$. Analogously when $l_n \asymp \Delta_n^{-(1/2+\delta)}$ we have the following estimates

TABLE 3
Estimates of ingredients under (3.12)

scaling properties	$E(\|\cdot\|^2\|\mathcal{F}_i^n)$	$\|E(\cdot\|\mathcal{F}_i^n)\|$
$R(p)_{i+v}'^n$	$p^2\Delta_n^{1+2\delta}$	$p\Delta_n^{1/2+\delta}$
$D(p)_{i+v}^n$	$p\Delta_n^{3/2-3\delta}$	$p\Delta_n^{1-2\delta}$
$A(p)_{i+v}^n$	$p^2\Delta_n^{2-2\delta}(p\Delta_n^{-1/2-\delta}+v)$	$p\Delta_n^{3/2-\delta}(p\Delta_n^{-1/2-\delta}+v)$
$\zeta(p)_{i,h}^n$	$p\Delta_n^{1-2\delta}$	$p^{3/2}\Delta_n^{1-2\delta}$

Let

$$\alpha(p)_{i,h}'^n = R(p+1)_{i+a(n,p,h)}'^n + D(p+1)_{i+a(n,p,h)}^n + A(p+1)_{i+a(n,p,h)}^n + \zeta(p+1)_{i,h}^n$$

Given $j,k,l,m=1,\cdots,d$, by Table 3 we have

$$\Big|E_i^n\big(\beta_i'^{n,jk}\beta_i'^{n,lm}\big) - (k_n\Delta_n^{1/2+\delta})^{-1}\Sigma(c_i^n)^{jk,lm}\Big| = \sum_{r=1}^5 \mu_i'^{n,r} + O_p\big(p(k_n^2\Delta_n^{1+2\delta})^{-1}\big)$$

where

$$\mu_i'^{n,1} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-1} E_i^n\left(\Big|\alpha(p)_{i,h}'^{n,jk}\alpha(p)_{i,h}'^{n,lm} - \zeta(p+1)_{i,h}^{n,jk}\zeta(p+1)_{i,h}^{n,lm}\Big|\right)$$

$$\mu_i'^{n,2} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-2}\sum_{h'=h+1}^{m(n,p)-1}\Big|E_i^n\left[\alpha(p)_{i,h}'^{n,jk}\alpha(p)_{i,h'}'^{n,lm} + \alpha(p)_{i,h}'^{n,lm}\alpha(p)_{i,h'}'^{n,jk}\right]\Big|$$

$$\mu_i'^{n,3} = \frac{1}{(k_n-l_n)^2\Delta_n^2}\sum_{h=0}^{m(n,p)-1} E_i^n\Bigg(\Big|\zeta(p+1)_{i,h}^{n,jk}\zeta(p+1)_{i,h}^{n,lm} - \Delta_n^{1/2-\delta}\frac{(p+1)\Phi_{00}-\Psi_{00}}{(p+2)\Phi_{00}}\Sigma\big(c_{i+a(n,p,h)}^n\big)^{jk,lm}(p+2)l_n\Delta_n\Big|\Bigg)$$

$$\mu_i'^{n,4} = \frac{[(p+1)\Phi_{00} - \Psi_{00}]\, l_n}{\Phi_{00}(k_n - l_n)^2 \Delta_n^{1/2+\delta}} \sum_{h=0}^{m(n,p)-1} \left| E_i^n \left[\Sigma\big(c^n_{i+a(n,p,h)}\big)^{jk,lm} - \Sigma(c_i^n)^{jk,lm} \right] \right|$$

$$\mu_i'^{n,5} = \frac{(p+1)\Phi_{00} - \Psi_{00}}{(p+1)\Phi_{00}} \left| \frac{(p+1)l_n}{(k_n - l_n)^2 \Delta_n^{1/2+\delta}} \left\lfloor \frac{k_n}{(p+1)l_n} \right\rfloor - \frac{1}{k_n \Delta_n^{1/2+\delta}} \right| \cdot \left| \Sigma(c_i^n)^{jk,lm} \right|$$
$$+ \frac{1}{k_n \Delta_n^{1/2+\delta}} \left| \frac{(p+1)\Phi_{00} - \Psi_{00}}{(p+1)\Phi_{00}} - 1 \right| \cdot \left| \Sigma(c_i^n)^{jk,lm} \right|$$

Using Table 3 to analyze $\mu_i'^{n,r}$ for $r = 1, \cdots, 5$ one by one, we get

$$\left| E_i^n \big(\beta_i'^{n,jk} \beta_i'^{n,lm}\big) - (k_n \Delta_n^{1/2+\delta})^{-1} \Sigma(c_i^n)^{jk,lm} \right| \le K \big[k_n \Delta_n + (p k_n \Delta_n^{1/2+\delta})^{-1} \big] \tag{D.8}$$

Define

$$\lambda'(x) = \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(x) \times \Sigma(x)^{jk,lm}$$
$$g_n'(x) = g(x) - (2k_n \Delta_n^{1/2+\delta})^{-1} \lambda'(x)$$

We have a decomposition analogous to (C.17):

$$\Delta_n^{-1/4+\delta/2} \big[\widetilde{S}(g)^n - S(g) \big] = \check{S}^{n,0} + \check{S}^{n,1} + \check{S}(p)^{n,2} + \check{S}^{n,3} + \check{S}(p)^{n,4} \tag{D.9}$$

where

$$\check{S}_t^{n,0} = \Delta_n^{-1/4+\delta/2} \left[\sum_{i=0}^{N_t^n - 1} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} g(c^n_{ik_n}) - g(c_s)\, \mathrm{d}s - \int_{N_t^n k_n \Delta_n}^{t} g(c_s)\, \mathrm{d}s \right]$$

$$\check{S}_t^{n,1} = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n - 1} \left[g_n'(\widetilde{c}^n_{ik_n}) - g_n'(\widetilde{c}^{*n}_{ik_n}) \right]$$

$$\check{S}(p)_t^{n,2} = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n - 1} \sum_{j,k=1}^{d} \partial_{jk} g(c^n_{ik_n}) \left[\xi_i^{n,0} + \xi_i^{n,1} + \xi_i'^{n,2} + N(p)^{n,jk}_{ik_n} \right]$$

$$\check{S}_t^{n,3} = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n - 1} \Big[g(\widetilde{c}^{*n}_{ik_n}) - g(c^n_{ik_n}) - \sum_{j,k=1}^{d} \partial_{jk} g(c^n_{ik_n}) \beta'^{n,jk}_{ik_n}$$
$$- (2k_n \Delta_n^{1/2+\delta})^{-1} \lambda'(\widetilde{c}_i^{*n}) \Big]$$

$$\check{S}(p)_t^{n,4} = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n - 1} \sum_{j,k=1}^{d} \partial_{jk} g(c^n_{ik_n}) \times M(p)^{n,jk}_{ik_n}$$

As in the proof of Theorem 1, by the argument for Lemma 4.4 in [21],

$$\check{S}^{n,0} \xrightarrow{u.c.p.} 0 \tag{D.10}$$

According to (C.19),

$$\|g_n'(x) - g_n'(y)\| \le K\|x - y\| + K(k_n \Delta_n^{1/2+\delta})^{-1} \|x - y\| \left(\|x\|^2 + \|x - y\|^2 \right)$$

and for large n, $\|g_n'(x) - g_n'(y)\| \leq K\|x-y\|$. By (D.2)

$$\mathbb{E}\left(\sup_{u\in[0,t]}\left\|\check{S}_u^{n,1}\right\|\right) \leq Kk_n\Delta_n^{3/4+\delta/2}\sum_{i=0}^{N_t^n-1}\left\|\widetilde{c}_{ik_n}^n - \widetilde{c}_{ik_n}^{*n}\right\|$$
$$\leq Kt\left[a_n\Delta_n^{(2-\nu)\rho+(1+2\delta)\nu/4-(3/4+\delta/2)} + \Delta_n^{3/4+\delta/2}\right]$$

Because of (3.12), $(2-\nu)\rho + (1+2\delta)\nu/4 - (3/4+\delta/2) > 0$, we have

$$\check{S}^{n,1} \xrightarrow{u.c.p.} 0 \tag{D.11}$$

Given $e_i^n \in \mathbb{R}^{d\times d}$, consider the process

$$\check{S}_t^{*n} = k_n\Delta_n^{3/4+\delta/2}\sum_{i=0}^{N_t^n-1}\sum_{j,k=1}^{d}\partial_{jk}g(c_{ik_n}^n)\times e_{ik_n}^{n,jk}$$

Assume the following holds

$$\begin{cases}\|E_i^n\left(e_i^n\right)\| & \leq K\Delta_n^{1/4-\delta/2}a_n\\ E_i^n\left(\|e_i^n\|^2\right) & \leq K(k_n\Delta_n^{1/2+\delta})^{-1}b_n\end{cases} \tag{D.12}$$

where $a_n, b_n \to 0$. Since ∂g is bounded by (B.1),

$$\mathbb{E}\left(\sup_{u\in[0,t]}\left\|\check{S}_u^{*n}\right\|\right) \leq Kk_n\Delta_n^{3/4+\delta/2}\sum_{i=0}^{N_t^n-1}\mathbb{E}\left(\left\|E_{ik_n}^n\left(e_{ik_n}^n\right)\right\|\right)$$
$$+Kk_n\Delta_n^{3/4+\delta/2}\mathbb{E}\left(\sup_{s\in[0,t]}\left\|\sum_{i=0}^{N_s^n-1}\left[e_{ik_n}^n - E_{ik_n}^n\left(e_{ik_n}^n\right)\right]\right\|\right)$$

note that $k_n\Delta_nN_t^n \asymp t$, by Lemma 1, $\mathbb{E}\left(\sup_{u\in[0,t]}\left\|\check{S}_u^{*n}\right\|\right) \leq K\left(ta_n + \sqrt{tb_n}\right) \to 0$.
We consider the following 4 cases,

① when $e_i^n = \xi_i^{n,0}$: by (C.4), (D.12) is satisfied with $a_n = k_n\Delta_n^{3/4+\delta/2}$, $b_n = (k_n\Delta_n^{3/4+\delta/2})^2$;
② when $e_i^n = \xi_i^{n,1}$: by (D.3), (D.12) is satisfied with $a_n = \Delta_n^{1/4-\delta/2}$, $b_n = k_n\Delta_n$;
③ when $e_i^n = \xi_i'^{n,2}$: by (D.4), (D.12) is satisfied with $a_n = \Delta_n^{5/(2\delta)-1/4}$, $b_n = k_n\Delta_n^{1/2+5\delta}$;
④ when $e_i^n = N(p)_i^n$: by (D.5), (D.12) is satisfied with $a_n = p^{-1}\Delta_n^{1/4-\delta/2}$, $b_n = p^{-1}$.

Therefore, provided we abide by the tuning choice (3.12) and let $p \asymp \Delta_n^{-1/12+\delta/6}$, we have

$$\check{S}(p)^{n,2} \xrightarrow{u.c.p.} 0 \tag{D.13}$$

Let $\iota_i'^n = \lambda'(\widehat{c}_i^{*n}) - \lambda'(c_i^n)$. By the definition (4.2) in the main article,

$$\|\iota_i'^n\| \leq K\left(\|\beta_i'^n\| + \|\beta_i'^n\|^2\right)$$

then by (3.12), (D.7),

$$E_i^n(\|\iota_i'^n\|) \leq Kk_n^{-1/2}\Delta_n^{-1/4-\delta/2} \tag{D.14}$$

We can rewrite $\check{S}^{n,3}$ as

$$\check{S}^{n,3} = G'^n + H'^n$$

where

$$G_t'^n = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n-1} \left[s_{ik_n}'^n + \frac{1}{2k_n \Delta_n^{1/2+\delta}} \iota_{ik_n}'^n + E_{ik_n}^n \left(v_{ik_n}'^n \right) \right]$$

$$H_t'^n = k_n \Delta_n^{3/4+\delta/2} \sum_{i=0}^{N_t^n-1} \left[v_{ik_n}'^n - E_{ik_n}^n \left(v_{ik_n}'^n \right) \right]$$

$$s_i'^n = g(c_i^n + \beta_i'^n) - g(c_i^n) - \sum_{j,k=1}^{d} \partial_{jk} g(c_i^n) \beta_i'^{n,jk} - \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(c_i^n) \beta_i'^{n,jk} \beta_i'^{n,lm}$$

$$v_i'^n = \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(c_i^n) \left[\beta_i'^{n,jk} \beta_i'^{n,lm} - (k_n \Delta_n^{1/2+\delta})^{-1} \Sigma(c_i^n)^{jk,lm} \right]$$

By (C.19), (D.7), (D.14), if we let $p \asymp \Delta_n^{-12+\delta/6}$,

$$\mathbb{E}\left(\sup_{s\in[0,t]} \|G_s'^n\| \right) \leq Kt \left[k_n \Delta_n^{3/4+\delta/2} + (k_n \Delta_n^{2(1+\delta)/3})^{-1} + \Delta_n^{25\delta/12-1/24} \right] \tag{D.15}$$

and

$$\begin{aligned} E_i^n(\|v_i'^n\|^2) &\leq K E_i^n \left[\|\beta_i'^n\|^4 + (k_n \Delta_n^{1/2+\delta})^{-1} \|\beta_i'^n\|^2 + (k_n \Delta_n^{1/2+\delta})^{-2} \right] \\ &\leq K \left[k_n \Delta_n + (k_n \Delta_n^{1/2+\delta})^{-2} \right] \end{aligned}$$

then Lemma 1 implies

$$\mathbb{E}\left(\sup_{s\in[0,t]} \|H_s'^n\| \right) \leq K\sqrt{t} \left[k_n \Delta_n^{3/4+\delta/2} + (k_n \Delta_n^{1/2+\delta})^{-1/2} \right] \tag{D.16}$$

Because of (3.12), (D.15), (D.16), we have

$$\check{S}^{n,3} \xrightarrow{u.c.p.} 0 \tag{D.17}$$

Recall the definition (C.26), rewrite $\check{S}(p)_t^{n,4}$ as

$$\check{S}(p)_t^{n,4} = \frac{k_n}{k_n - l_n} \sum_{j,k=1}^{d} \Delta_n^{-1/4+\delta/2} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \Lambda(p)_{i,h}^{n,jk}$$

By the last line of Table 3 and Jensen's inequality,

$$\Delta_n^{-1/4+\delta/2} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \left\| E\left[\Lambda(p)_{i,h}^n | \mathcal{H}(p)_{i,h}^n \right] \right\| \leq K p^{1/2} \Delta_n^{1/4-\delta} \tag{D.18}$$

$$\Delta_n^{-1+2\delta} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} E\left[\left\| \Lambda(p)_{i,h}^n \right\|^4 | \mathcal{H}(p)_{i,h}^n \right] \leq K p \Delta_n^{1/2-\delta} \tag{D.19}$$

The same argument resulting in (5.58) in [12] gives us

$$\Delta_n^{-1/4+\delta/2} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} \left\| E\left[\Lambda(p)_{i,h}^n \Delta N(p)_{i,h}^n | \mathcal{H}(p)_{i,h}^n \right] \right\| \xrightarrow{\mathbb{P}} 0 \tag{D.20}$$

where $\Delta N(p)^n_{i,h} = N^n_{ik_n+b(n,p,h)} - N^n_{ik_n+a(n,p,h)}$ and N is a bounded martingale orthogonal to W or $N = W^l$ for some $l = 1, \cdots, d'$.
Calculation similar to that of $\mu_i'^{n,3}$ shows

$$\text{(D.21)} \quad \Delta_n^{-1/2+\delta} \sum_{i=0}^{N_t^n-1} \sum_{h=0}^{m(n,p)-1} E\left[\Lambda(p)^{n,jk}_{i,h}\,\Lambda(p)^{n,lm}_{i,h} | \mathcal{H}(p)^n_{i,h}\right] = \sum_{i=0}^{N_t^n-1} \partial_{jk} g(c^n_{ik_n}) \partial_{lm} g(c^n_{ik_n}) \times \sum_{h=0}^{m(n,p)-1} \frac{p\Phi_{00} - \Psi_{00}}{(p+1)\Phi_{00}} \Sigma\big(c^n_{ik_n+a(n,p,h)}\big)^{jk,lm} (p+1) l_n \Delta_n + t p^{3/2} O_p(\Delta_n^{1/4-\delta/2}) \xrightarrow{\mathbb{P}} \int_0^t \partial_{jk} g(c_s) \partial_{lm} g(c_s)^{\mathrm{T}}\, \Sigma(c_s;p)^{jk,lm}\, \mathrm{d}s$$

To establish stable convergence, we then only need to combine (D.18), (D.19), (D.20), (D.21) and apply Theorem IX.7.28 in [34]. We get for $\forall p \in \mathbb{N}^+$,

$$\text{(D.22)} \qquad \check{S}(p)^{n,4} \xrightarrow{\mathcal{L}-s(f)} Z(p)$$

where $Z(p)$ is a process defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, such that conditioning on $\mathcal{F}$ it is a mean-0 continuous Itô martingale with variance

$$\widetilde{E}[Z(p)Z(p)^{\mathrm{T}}|\mathcal{F}] = \int_0^t \sum_{j,k,l,m=1}^{d} \partial_{jk} g(c_s) \partial_{lm} g(c_s)^{\mathrm{T}} \frac{p\Phi_{00} - \Psi_{00}}{(p+1)\Phi_{00}} \Sigma(c_s)^{jk,lm}\, \mathrm{d}s$$

where $\widetilde{E}$ is the conditional expectation operator on the extended probability space.
By (C.20), (D.11), (D.13), (D.17), (D.22), as $p \to \infty$, we arrive at the asymptotic result of Theorem 2 in the main article.

APPENDIX E: DERIVATION FOR PROPOSITION 3 AND 4

In the following lemmas and derivations, we suppress the time index in the subscripts to reduce notational clutter. From now on, we write $c = c_t$, $\lambda^r = \lambda^r_t$, $q^r = q^r_t$ is the eigenvector associated with eigenvalue λ^r, $q^r_k = q^r_{t,k}$ is the k-th entry in q^r, and the eigenvalue factorization is written as

$$\text{(E.1)} \qquad c = \left[q^1, \cdots, q^d\right] \begin{bmatrix} \lambda^1 & & \\ & \ddots & \\ & & \lambda^d \end{bmatrix} \begin{bmatrix} q^{1,\mathrm{T}} \\ \vdots \\ q^{d,\mathrm{T}} \end{bmatrix}$$

where $\lambda_1 \geq \cdots \geq \lambda_d$.
The following lemma is a restatement of Lemma 2 in [4].

LEMMA 10. *When λ^r is a simple eigenvalue in (E.1), the functions $\lambda^r(c)$ and $q^r(c)$ are twice continuously differentiable. We have the following result about gradients:*

$$\partial_{jk}\lambda^r(c) = q^r_j\, q^r_k$$

$$\partial_{jk} q^r(c) = \left(\lambda^r\, \mathbf{I}_d - c\right)^{\dagger} \times \partial_{jk} c \times q^r$$

where the superscript notation † refers to Moore-Penrose inverse, equivalently,

$$\partial_{jk} q^r(c) = \sum_{v \neq r} \frac{q_j^v \, q_k^r}{\lambda^r - \lambda^v} \times q^v$$

Moreover, the Hessians are

$$\partial^2_{jk,lm} \lambda^r(c) = \sum_{v \neq r} \frac{q_j^v \, q_l^v \, q_k^r \, q_m^r + q_k^v \, q_l^v \, q_j^r \, q_m^r}{\lambda^r - \lambda^v}$$

and

$$\partial^2_{jk,lm} q^r(c) = \sum_{v \neq r} \frac{q_l^v \, q_m^v \, q_j^v \, q_k^r - q_j^v \, q_k^r \, q_l^r \, q_m^r}{(\lambda^r - \lambda^v)^2} \times q^v + \sum_{v \neq r} \sum_{b \neq v} \frac{q_l^b \, q_m^v \, q_j^v \, q_k^r}{(\lambda^r - \lambda^v)(\lambda^v - \lambda^b)} \times q^b$$
$$+ \sum_{v \neq r} \sum_{b \neq v} \frac{q_j^b \, q_l^b \, q_m^v \, q_k^r}{(\lambda^r - \lambda^v)(\lambda^v - \lambda^b)} \times q^v + \sum_{v \neq r} \sum_{b \neq r} \frac{q_k^b \, q_l^b \, q_m^r \, q_j^v}{(\lambda^r - \lambda^v)(\lambda^r - \lambda^b)} \times q^v$$

The following lemmas is a direct consequence Lemma 4 in [4].

LEMMA 11. *Suppose the eigenvalues of c form K clusters in the sense that*

$$\lambda^{r_0+1} = \cdots = \lambda^{r_1} > \lambda^{r_1+1} = \cdots = \lambda^{r_2} > \cdots\cdots > \lambda^{r_{K-1}+1} = \cdots = \lambda^{r_K}$$

where $r_0 = 0$, $r_K = d$. We define the following functions on $\mathbb{R}^d$

$$f_h(x) = \frac{1}{r_h - r_{h-1}} \sum_{j=r_{h-1}+1}^{r_h} x^j, \; h = 1, \cdots, K$$

Define function $\lambda(c) = (\lambda^1, \cdots, \lambda^d)$, and functions $F_h = f_h \circ \lambda$, $h = 1, \cdots, K$, where $\circ$ denotes function composition. Then we have

$$\partial_{jk} F_h(c) = \frac{1}{r_h - r_{h-1}} \sum_{r=r_{h-1}+1}^{r_h} q_j^r \, q_k^r$$

$$\partial^2_{jk,lm} F_h(c) = \frac{1}{r_h - r_{h-1}} \sum_{r=r_{h-1}+1}^{r_h} \left[\left(\lambda^r \mathbf{I}_d - c\right)^{\dagger, km} q_j^r \, q_l^r + \left(\lambda^r \mathbf{I}_d - c\right)^{\dagger, jl} q_k^r \, q_m^r \right]$$

where the superscript notation † stands for Moore-Penrose inverse, equivalently,

$$\partial^2_{jk,lm} F_h(c) = \frac{1}{r_h - r_{h-1}} \sum_{r=r_{h-1}+1}^{r_h} \sum_{\lambda^v \neq \lambda^r} \frac{q_j^r \, q_l^r \, q_k^v \, q_m^v + q_j^v \, q_l^v \, q_k^r \, q_m^r}{\lambda^r - \lambda^v}$$

REMARK 3. If $r_h = r_{h-1} + 1$, $F_h(c) = \lambda^{r_h}(c)$. From Lemma 10, 11 we see that $\lambda^r(c)$ is first-order differentiable regardless of λ^r being a simple or repeated eigenvalue of c. However, when λ^r is a repeated eigenvalue, $\lambda^r(c)$ is not second-order differentiable; as for the eigenvector, $q^r(c)$ is not differentiable at all.

By some calculation, we can see that Proposition 3 and 4 are consequences of Theorem 2 in the main article and Lemma 10, 11.

APPENDIX F: SIMULATIONS

F.1. Simulation of the rate-optimal CLT. We simulate the following noisy observations of scalar jump-diffusion with stochastic volatility via the Euler scheme:

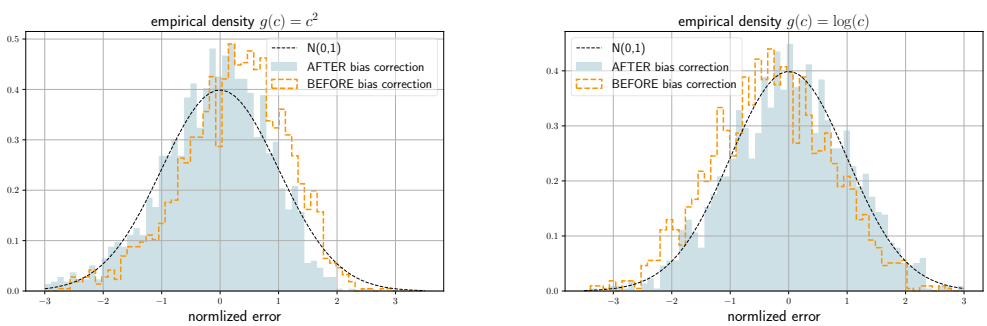
$$\begin{cases} Y_i^n &= X_i^n + \varepsilon_i^n \\ \mathrm{d}X_t &= .03\,\mathrm{d}t + \sqrt{c_t}\,\mathrm{d}W_t + J_t^X\,\mathrm{d}N_t^X \\ \mathrm{d}c_t &= 6(.16 - c_t)\,\mathrm{d}t + .5\sqrt{c_t}\,\mathrm{d}B_t + \sqrt{c_{t-}}J_t^c\,\mathrm{d}N_t^c \end{cases}$$

where $\varepsilon_i^n \overset{\text{i.i.d.}}{\sim} N(0, .005^2)$, $\mathbb{E}[(W_{t+\Delta} - W_t)(B_{t+\Delta} - B_t)] = -.6\Delta$, $J_t^X \sim N(-.01, .02^2)$, $N_{t+\Delta}^X - N_t^X \sim \mathrm{Poisson}(36\Delta)$, $\log(J_t^c) \sim N(-5, .8)$, $N_{t+\Delta}^c - N_t^c \sim \mathrm{Poisson}(12\Delta)$.
Each simulation employs 23400×21 data points with $\Delta_n = 1$ second. We choose the following tuning parameters:

functionals	l_n	k_n	ν_n
$g(c) = c^2$	$\lfloor \Delta_n^{-.5} \rfloor$	$\lfloor \Delta_n^{-.69} \rfloor$	$1.6\overline{\sigma}^2\Delta_n^{.47}$
$g(c) = \log(c)$	$\lfloor \Delta_n^{-.5} \rfloor$	$\lfloor \Delta_n^{-.7} \rfloor$	$1.5\overline{\sigma}^2\Delta_n^{.47}$

where $\overline{\sigma}^2$ is an estimate of the average volatility by bipower variation [35].
We compute estimation errors in the simulation, and normalize the errors by the corresponding asymptotic variances. The results are shown in Figure 7.

FIG 7. *Simulation of estimators of scalar volatility functionals*



F.2. Simulation of PCA of a factor model. We simulate the following factor model:

$$Y_t = X_t + e_t$$
$$X_t = X_0 + \int_0^t \beta_s\,\mathrm{d}F_s + Z_t,$$

where $\beta_s \in \mathbb{R}^{d\times r}$ is the factor loading matrix with $r \le d$, $F_s \in \mathbb{R}^r$ is the latent factor, $Z_t \in \mathbb{R}^d$ is idiosyncratic component satisfying $Z \perp\!\!\!\perp F$, and e_t is a multidimensional noise process exhibits no autocorrelation but is possibly cross-sectionally correlated.
In the component-wise differential form, we can write X_t as

$$\mathrm{d}X_t^j = \sum_{k=1}^r \beta_t^{jk}\,\mathrm{d}F_t^k + \mathrm{d}Z_t^j$$

$$\mathrm{d}\beta_t^{j1} = \kappa^{j1}\big(\theta^{j1} - \beta_s^{j1}\big)\,\mathrm{d}s + \xi^{j1}\sqrt{\beta_t^{j1}}\,\mathrm{d}\breve{W}_t^{j1}$$

$$\mathrm{d}\beta_t^{jk} = \kappa^{jk}\big(\theta^{jk} - \beta_s^{jk}\big)\,\mathrm{d}s + \xi^{jk}\,\mathrm{d}\breve{W}_t^{jk},\ k \geq 2$$
$$\mathrm{d}F_t^k = \mu^k\,\mathrm{d}t + \sqrt{\Pi_t^k}\,\mathrm{d}W_t^k + J_t^{F,k}\,\mathrm{d}N_t^{F,k}$$
$$\mathrm{d}Z_t^j = \chi_t\,\mathrm{d}B_t^j + J_t^{Z,j}\,\mathrm{d}N_t^{Z,j},$$

and

$$\mathrm{d}\Pi_t^k = \widetilde{\kappa}^k\big(\widetilde{\theta}^k - \Pi_t^k\big)\,\mathrm{d}t + \eta^k\sqrt{\Pi_t^k}\,\mathrm{d}\widetilde{W}_t^k + J_t^{\Pi,k}\,\mathrm{d}N_t^{F,k}$$
$$\mathrm{d}\chi_t^2 = \kappa\big(\theta - \chi_t^2\big)\,\mathrm{d}t + \eta\chi_t\,\mathrm{d}\widetilde{B}_t,$$

where W and $\widetilde{W}$ are $\mathbb{R}^r$-valued standard Brownian motions, $\breve{W}$ is a $\mathbb{R}^{d\times r}$-valued Brownian motion with independent elements, B is a $\mathbb{R}^d$-valued standard Brownian motion, $\widetilde{B}$ is a scalar standard Brownian motion. Except W and $\widetilde{W}$, all Brownian motions in this model are mutually independent.

In this simulation model, the factor loading, factor volatility and idiosyncratic volatility are modeled by Ornstein-Uhlenbeck processes so that they exhibit mean-reverting dynamics. Especially, the first factor is presumably the market factor, hence the associated factor loading β_t^{j1} $j = 1, \cdots, d$ are positive almost surely.

Moreover, for $k = 1, \cdots, r$, $\mathbb{E}\big[(W_{t+\Delta}^k - W_t^k)(\widetilde{W}_{t+\Delta}^k - \widetilde{W}_t^k)\big] = \rho^k\Delta$ is the leverage effect in the k-th factor; $N_{t+\Delta}^{F,k} - N_t^{F,k} \sim \mathrm{Poisson}(\lambda^{F,k}\Delta)$ is the count of co-jumps of the factor and factor volatility, $J_t^{F,k} \overset{\text{i.i.d.}}{\sim} \mathrm{Laplace}(\zeta^{F,k})$ is the size of factor jump, $J_t^{\Pi,k} \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\zeta^{\Pi,k})$ is the size of volatility jump; for $j = 1, \cdots, d$, $N_{t+\Delta}^{Z,j} - N_t^{Z,j} \sim \mathrm{Poisson}(\lambda^{Z,j}\Delta)$ is the count of idiosyncratic jumps, $J_t^{Z,j} \overset{\text{i.i.d.}}{\sim} \mathrm{Laplace}(\zeta^{Z,j})$ is the size of idiosyncratic jump.

We set $d = 30$, $r = 3$, and generate 22800×21 data points with $\Delta_n = 1$ second in each simulation. We choose the following tuning parameters in PCA:

functionals	l_n	k_n	ν_n
$g(c) = \lambda^1(c)$	$\lfloor .23\Delta_n^{-.62}\rfloor$	$\lfloor .57\Delta_n^{-.75}\rfloor$	$1.5\overline{c}\Delta_n^{.47}$
$g(c) = \lambda^2(c)$	$\lfloor .038\Delta_n^{-.62}\rfloor$	$\lfloor .19\Delta_n^{-.75}\rfloor$	$1.5\overline{c}\Delta_n^{.47}$
$g(c) = q^{1,1}(c)$	$\lfloor .23\Delta_n^{-.62}\rfloor$	$\lfloor .57\Delta_n^{-.75}\rfloor$	$1.5\overline{c}\Delta_n^{.47}$
$g(c) = q^{1,2}(c)$	$\lfloor .23\Delta_n^{-.62}\rfloor$	$\lfloor .57\Delta_n^{-.75}\rfloor$	$1.5\overline{c}\Delta_n^{.47}$

where $\overline{c}$ is the vector of time-averaged estimates of volatility matrix diagonal elements computed by bipower variations. We conduct element-wise jump truncation by the rule $|\overline{Y}_i^{r,n}| > 1.5\overline{c}^r\Delta_n^{.47}$.

Figure 8 and Figure 9 show empirical densities of studentized estimation errors.

FIG 8. *Simulation of eigenvalue estimators*

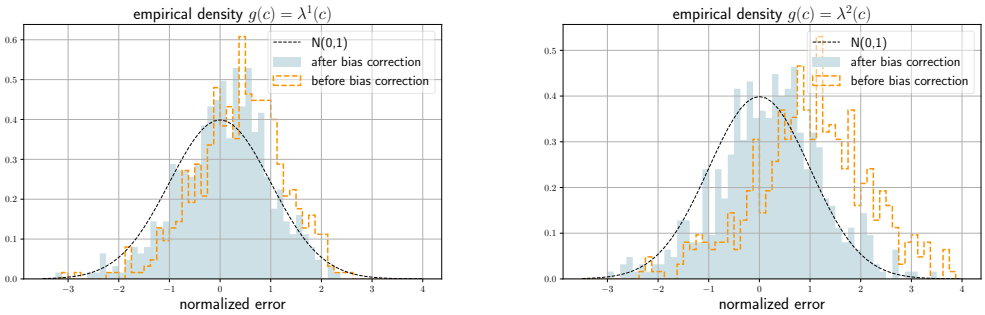
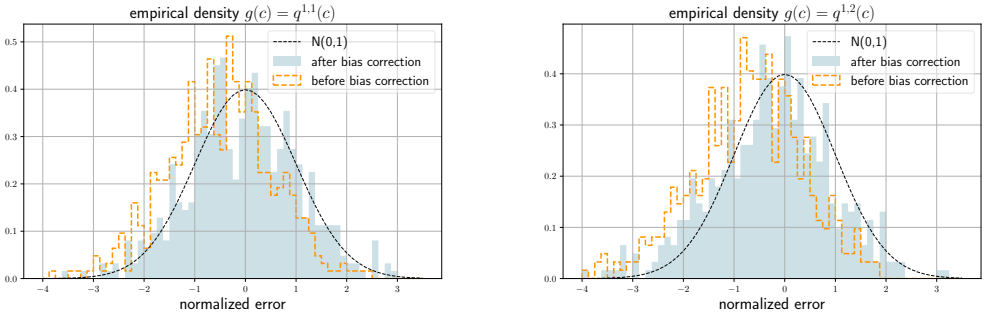


FIG 9. *Simulation of eigenvector estimators*



REFERENCES

[1] TODOROV, V., TAUCHEN, G. (2012). The realized Laplace transform of volatility. *Econometrica* 80 (3), 1105-1127.

[2] LI, J., XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* 84 (4), 1613-1633.

[3] LI, J., TODOROV, V., TAUCHEN, G. (2017). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics* 200, 36-47.

[4] AÏT-SAHALIA, Y., XIU, D. (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114 (525), 287-303.

[5] LI, J., LIU, Y., XIU, D. (2019). Efficient estimation of integrated volatility functionals via multiscale Jackknife. *The Annals of Statistics* 47 (1), 156-176.

[6] YANG, X. (2021). Semiparametric estimation in continuous-time: asymptotics for integrated volatility functionals with small and large bandwidths. *Journal of Business & Economic Statistics* 39 (3), 793-806.

[7] ZHANG, L., MYKLAND, P., AÏT-SAHALIA, Y. (2005). A tale to two scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100 (472), 1394-1411.

[8] ZHANG, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160, 33-47.

[9] BARNDORFF-NIELSEN, O.E., HANSEN, P.R., LUNDE, A., SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76 (6), 1481-1536.

[10] BARNDORFF-NIELSEN, O.E., HANSEN, P.R., LUNDE, A., SHEPHARD, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149-169.

[11] PODOLSKIJ, M., VETTER, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15 (3), 634-658.

[12] JACOD, J., LI, Y., MYKLAND, P., PODOLSKIJ, M., VETTER, M. (2009). Microstructure noise in the continuous case: the preaveraging approach. *Stochastic processes and their applications* 119, 2249-2276.

[13] JACOD, J., PODOLSKIJ, M., VETTER, M. (2010). Limit theorems for moving averages of discretized processes plus noise. *The Annals of Statistics* 38 (3), 1478-1545.

[14] CHRISTENSEN, K., KINNEBROCK, S., PODOLSKIJ, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* 159, 116-133.

[15] XIU, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235-250.

[16] AÏT-SAHALIA, Y., FAN, J., XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105 (492), 1504-1517.

[17] SHEPHARD, N., XIU, D. (2017). Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading. *Journal of Econometrics* 201, 19-42.

[18] BIBINGER, M., HAUTSCH, N., MALEC, P., REIß, M. (2014). Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. *The Annals of Statistics* 42 (4), 1312-1346.

[19] BIBINGER, M., HAUTSCH, N., MALEC, P., REIß, M. (2019). Estimating the spot covariation of asset prices - statistical theory and empirical evidence. *Journal of Business & Economic Statistics* 37 (3), 419-435.

[20] MYKLAND, P., ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* 77 (5), 1403-1445.

[21] JACOD, J., ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: efficient estimation. *The Annals of Statistics* 41 (3), 1462-1484.

[22] CHEN, R. (2019). The Fourier transform method for volatility functional inference by asynchronous observations. arXiv.1911.02205.

[23] TODOROV, V., TAUCHEN, G., GRYNKIV, I. (2011). Realized Laplace transforms for estimation of jump diffusive volatility models. *Journal of Econometrics* 164, 367-381.

[24] TODOROV, V., TAUCHEN, G. (2012). Inverse realized Laplace transforms for nonparametric volatility density estimation in jump-diffusions. *Journal of the American Statistical Association* 107 (498), 622-635.

[25] LI, J., TODOROV, V., TAUCHEN, G. (2016). Inference theory for volatility functional dependencies. *Journal of Econometrics* 193, 17-34.

[26] YANG, X. (2020). Time-invariant restrictions of volatility functionals: efficient estimation and specification tests. *Journal of Econometrics* 215, 486-516.

[27] CHEN, D., MYKLAND, P., ZHANG, L. (2020). The five trolls under the bridge: principal component analysis with asynchronous and noisy high frequency data. *Journal of the American Statistical Association* 115 (532), 1960-1977.

[28] JACOD, J., ROSENBAUM, M. (2015). Estimation of volatility functionals: the case of a $\sqrt{n}$ window. In Friz, P. et al. *Large Deviations and Asymptotic Methods in Finance*, Springer.

[29] LI, J. (2013). Robust estimation and inference for jumps in noisy high frequency data: a local-to-continuity theory for the pre-averaging method. *Econometrica* 81 (4), 1673-1693.

[30] MYKLAND, P., ZHANG, L. (2016). Between data cleaning and inference: pre-averaging and robust estimators of the efficient price. *Journal of Econometrics* 194, 242-262.

[31] AÏT-SAHALIA, Y., XIU, D. (2019). A Hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics* 211, 176-205.

[32] PELGER, M. (2020). Understanding systematic risk: a high-frequency approach. *Journal of Finance* Volume 75, Issue 4.

[33] JACOD, J., PROTTER, P. (2012). *Discretization of Processes*. Springer-Verlag Berlin Heidelberg.

[34] JACOD, J., SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2ed. Springer-Verlag Berlin Heidelberg.

[35] PODOLSKIJ, M., VETTER, M. (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications* 119, 2803-2831.