Spot Variance Regressions *

Jia Li
Department of Economics
Duke University

Dacheng Xiu
Booth School of Business
University of Chicago

This Version: December 5, 2012
(very preliminary and incomplete)

Abstract

We study a nonlinear vector regression model for discretely sampled high-frequency data with the latent stochastic volatility as an explanatory variable. We propose a two-stage inference procedure by first nonparametrically recovering the volatility path from asset returns and then conducting inference based on the generalized method of moments (GMM). The GMM estimator is nonstandard in that the second-order asymptotics is dominated by a bias term, rendering standard inference implausible. We propose several bias-correction methods and show that the bias-corrected estimators have parametric rate of convergence with mixture normal distributions. Tests for overidentification and parameter stability are provided, followed by Andersen-Rubin type confidence sets that are robust to weak identification.

The results of the paper are applied to a linear model and a possibly weakly identified nonlinear model to the VIX index. Substantive evidence is obtained against the conventional models in the literature.

Keywords: high frequency data; semimartingale; VIX; spot volatility; bias correction; weak identification.

JEL Codes: C22.

*We are grateful to Tim Bollerslev, George Tauchen, ... for helpful comments.

1 Introduction

We consider the estimation and inference of a continuous-time vector regression model involving the latent volatility of a financial asset:

$$Y_t = f\left(Z_t, V_t, ; \theta\right) + \varepsilon_t, \tag{1.1}$$

where $\theta \in \Theta$ is the finite-dimensional parameter of interest, V_t is the latent spot variance of an asset price process X_t, and (Y_t, X_t, Z_t) are observed at discrete times. Such models are commonly used in the empirical option pricing literature, where Y_t is the market price of one or several options, $f(\cdot; \theta)$ is a theoretical pricing function with known functional form up to the parameter θ, X_t is the price of the underlying asset, Z_t includes observable state variables such as time, the risk-free rate and often X_t, and ε_t is the pricing error.

Direct estimation and inference based on model (1.1) is infeasible because V_t is unobservable. Nevertheless, information regarding V_t can be extracted from the asset price X_t. We consider a general setting where X_t follows an Itô semimartingale given by

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_t dW_t + J_t, \tag{1.2}$$

where b_t is the instantaneous drift, $\sigma_t = \sqrt{V_t}$ is the stochastic volatility, W_t is a Brownian motion and J_t is a jump process. The Itô semimartingale model is quite general and includes most continuous-time models in economics and finance. Inferring V_t from X_t in general is quite complicated as the volatility is not only convoluted with the Brownian shocks but also the drift and jump components of the price process. The standard practice in the empirical option pricing literature is to augment the pricing model (1.1) with a parametric version of (1.2), often tightly parametrized to maintain computational feasibility, and then conduct joint inference. Such procedure may result in misleading inference concerning the pricing model (1.1) when the auxiliary model is misspecified.

This paper proposes an alternative solution to the latent volatility problem by using intraday data sampled at high frequency. When X is sampled discretely but with mesh of the observation grid shrinking to zero, i.e., high-frequency data are available, one can nonparametrically recover the volatility path (see e.g. chapter 9 in Jacod and Protter (2012)). A natural estimation procedure, which we propose here, is to conduct estimation of (1.1) via the generalized method of moments (GMM) but with V_t replaced with its nonparametric estimate; we henceforth refer to this estimator as the GMM estimator. This procedure is different from a standard two-stage GMM estimation, because the first-stage estimation here is nonparametric in nature, but not kernel- or sieve-based, and constructed from (possibly discontinuous) semimartingales instead of independent or weakly dependent data as is typical in cross-sectional or standard time-series settings.

We show that the GMM estimator is consistent under the fill-in asymptotics with fixed time span and asymptotically vanishing sampling interval of the high-frequency data. A striking result of this paper is that the GMM estimator, when centered at the true parameter value, does not achieve a central limit theorem (CLT). Instead, the second-order asymptotic behavior of the GMM estimator is dominated by a bias term arising from the first-stage estimation of the spot variance, rendering standard inference implausible. This phenomenon is in sharp contrast with the standard two-stage GMM setting, where a CLT is available and the first-stage estimation only affects the asymptotic variance.

We explicitly characterize the higher-order bias in the "raw" GMM estimator and propose a bias-correction. We show that the bias-corrected estimator, henceforth the two-step estimator, is centered at the true parameter and admits a CLT with parametric rate of convergence. We also propose estimators for the asymptotic variance in explicit forms for the purpose of inference. Efficient GMM estimation can then be implemented by optimally choosing the weighting matrix. The bias-correction is simple to implement and performs well in simulations. We also propose a one-step estimator that is based on a bias-corrected GMM objective function and automatically conducts bias-correction within the estimation step. We show that the one-step estimator is second-order asymptotically equivalent to the two-step estimator. We further propose a continuous updating estimator in the sense of Hansen et al. (1996), which achieves both bias-correction and efficient choice of weighting matrix within one step. We show that the continuous updating estimator is second-order asymptotically equivalent to the efficient two-step and one-step estimators.

We further propose tools for inference. First, we construct overidentification tests for misspecification. Second, in a multiperiod setting, in which the estimation is conducted period by period, we examine the parameter stability across periods. Formally, we construct uniform confidence band for the path of each component of the true parameter. We also provide a general class of tests for the hypothesis of parameter stability. Finally, we construct Anderson-Rubin type confidence sets for the true parameter, which is robust to weak identification.

The results of the paper are applied to a linear model for the VIX index, where the linear specification is implied by a large class of structural models. We document strong evidence for parameter instability. We further generalize the linear model to a nonlinear model, which we label as the "ABC" model, that is possibly weakly identified. We show that the ABC model serves as a parsimonious approximation for a more complicated structural VIX pricing model based on exponential-Levy risk-neutral volatility dynamics. When nested in the ABC model, the linear specification is strongly rejected during the financial crisis (2008 Q3-Q4), the European debt crisis (2010 Q2) and the U.S. debt crisis (2011 Q3), but not other periods in our sample (2007 Q1 to 2012 Q3).

Our main contribution is twofold. First, we extend the high-frequency financial econometrics literature into a general GMM framework with the spot variance as a regressor. We demonstrate the failure of the naive approach in which one ignores the error in the estimation of spot volatility. Constructively, we provide a comprehensive econometric toolkit for bias-correction and inference. Second, our method offers a new empirical framework, especially for empirical option pricing. The framework solves the latent volatility problem by nonparametrically extracting the volatility path in an essentially model-free way, allowing for general volatility dynamics, such as leverage effect and volatility jumps with infinite activity or even infinite variation. The framework readily accommodates most stochastic volatility models in finance. As both the availability and the quality of high-frequency datasets have increased rapidly in recent years, we expect this new econometric framework to provide new insights by taking advantage of the richer information of such datasets. We demonstrate the usefulness of these tools with a novel empirical application on VIX pricing.

We now discuss the related literature. The current paper is closely related to the literature on nonparametric inference for volatility functionals; see Andersen et al. (2003), Barndorff-Nielsen et al. (2008), Gonçalves and Meddahi (2009), Todorov and Tauchen (2011) and, in particular, the recent work of Jacod and Rosenbaum (2012). Our focus here is very different from these papers. While the aforementioned papers focus on the inference of the volatility process itself, we treat the estimation of volatility only as a preliminary step and mainly consider the inference of structural parameters in economic models with the stochastic volatility as an explanatory variable. The current paper is also related to the study of realized beta and leverage effect (Mykland and Zhang (2009), Todorov and Bollerslev (2010)), which can be interpreted as linear regressions of high-frequency returns of an asset on the return of a risk factor. Here, we consider general nonlinear models on the level, instead of return, of economic and financial variables, in a GMM setting.

The paper is clearly related to the vast literature on GMM, in particular Hansen (1982) and Hansen et al. (1996). The Anderson-Rubin type confidence sets proposed in Section 3.3 are similar to those of Stock and Wright (2000) and Andrews and Soares (2010). This said, our theory is very different than the existing literature due to the complication of the first-stage nonparameteric recovery of the stochastic volatility path from the price returns, hence also the analysis in the second-stage and subsequent inference. The second-order bias induced by the first-stage estimation of the volatility path, as well as the bias-correction, appear to be a unique phenomenon in the high-frequency setting considered here. Moreover, the fill-in asymptotic setting considered here allows for considerable dependence and heterogeneity in the underlying processes such as the asset price and its stochastic volatility. This feature is in sharp contrast to the "large T" asymptotics in standard time-series analysis, which typically requires high-level conditions on weak dependence and moderate heterogeneity.

The paper is organized as follows. Section 2 presents the main theory. Section 3 presents tests for overidentification and parameter stability, as well as confidence sets that are robust to weak identification. Section 4 shows simulation results, followed by an empirical application in 5. Section 6 concludes. All proofs are in the appendix.

All limits below are taken as $n \to \infty$. All vectors are column vectors. The transpose of any matrix A is denoted by $A^\intercal$. For notational simplicity, we write (a, b) instead of $(a^\intercal, b^\intercal)^\intercal$. For any vector x in some finite-dimensional space $\mathcal{X}$, we use $\dim(x)$ and $\dim(\mathcal{X})$ interchangeably to denote the dimensionality of $\mathcal{X}$. We use $\|\cdot\|$ to denote the Euclidean norm. For any $\mathbb{R}^q$-valued function $f(x, y)$, $x \in \mathbb{R}^{\dim(x)}$, $y \in \mathbb{R}$, we denote by $\partial_x f(x, y)^\intercal$ its $\dim(x) \times q$ partial derivative matrix w.r.t. x, with the transpose sign suppressed when $q = 1$. Similarly, $\partial_x^2 f(x, y)$ denotes the $\dim(x) \times \dim(x)$ Hessian matrix of $f(x, y)$ w.r.t. x and $\partial_y^j f(x, y)$ is the jth partial derivative of $f(x, y)$ w.r.t. y for $j \geq 0$.

2 The main theory

2.1 The raw estimator and its consistency

We consider a multiple equation GMM setup. For $1 \leq j \leq J$,

$$Y_{j,t} = Y_{j,t}^* + \varepsilon_{j,t}, \quad Y_{j,t}^* = f_j(Z_t, V_t; \theta_0),$$

where $Y_{j,t}$, Z_t and V_t take values in $\mathcal{Y} \subseteq \mathbb{R}$, $\mathcal{Z} \subseteq \mathbb{R}^{\dim(\mathcal{Z})}$ and $\mathcal{V} \subseteq \mathbb{R}_+$ respectively, $\theta_0 \in \Theta \subseteq \mathbb{R}^{\dim(\Theta)}$ is the true parameter and the parameter space Θ is compact. The latent spot variance V_t is linked with the price X_t by (1.2). Regularity conditions are given below. We suppose that the data (Y_t, Z_t, X_t) are observed at discrete times $i\Delta_n$, $i = 0, 1, \ldots$ on $[0, T]$ for a fixed $T > 0$ with the time lag $\Delta_n \to 0$. For each $i \geq 0$, we denote $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

We approximate the latent spot variance $V_{i\Delta_n}$ with a local estimator $\widehat{V}_{i\Delta_n}$. To this end, we consider a sequence of integerss k_n with $k_n \to \infty$ and $k_n\Delta_n \to 0$. For $i = 0, \ldots, [T/\Delta_n] - k_n$, we set

$$\widehat{V}_{i\Delta_n} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right)^2 1_{\{|\Delta_{i+j}^n X| \leq u_n\}}, \tag{2.1}$$

where the truncation level u_n satisfies

$$u_n = \bar{\alpha} \Delta_n^{\varpi}, \quad \bar{\alpha} > 0 \quad \text{and} \quad \varpi \in (0, 1/2).$$

The truncation indicator in (2.1) ensures that the local approximation is robust to jumps in X. See Chapter 9 of Jacod and Protter (2012) for more discussions.

For equation j, we consider a $k_j \times 1$ instrument $d_j(Z_t, V_t; \theta)$ and set $g_j(y, z, v; \theta) = d_j(z, v; \theta)(f_j(z, v; \theta) - y_j)$. We denote $g(\cdot) = (g_1(\cdot), \ldots, g_J(\cdot))$. The collection of $k \equiv \sum_{j=1}^J k_j$

sample moment functions is given by

$$\widehat{G}_n(\theta) = \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} g\left(Y_{i\Delta_n}, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right).$$

For some weighting matrix Ξ_n converging in probability to a $\mathcal{F}$-measurable random matrix Ξ, we define the GMM estimator as follows:

$$\hat{\theta}_n \equiv \arg\min_{\theta\in\Theta} \widehat{Q}_n(\theta), \quad \text{where} \quad \widehat{Q}_n(\theta) \equiv \widehat{G}_n(\theta)^\intercal \Xi_n \widehat{G}_n(\theta). \tag{2.2}$$

Under regularity conditions, we show (see Theorem 1 below)

$$\widehat{G}_n(\theta) \xrightarrow{\mathbb{P}} G(\theta), \quad \widehat{Q}_n(\theta) \xrightarrow{\mathbb{P}} Q(\theta) \tag{2.3}$$

where

$$G(\theta) \equiv \int_0^T g(Y_s^*, Z_s, V_s; \theta)\, ds, \quad Q(\theta) \equiv G(\theta)^\intercal \Xi G(\theta).$$

The GMM problem considered here is therefore nonstandard because the limit, or "population", moment function $G(\cdot)$ is not expressed in terms of expectations like in the standard GMM setting, but rather defined as an integrated stochastic quantity over $[0,T]$. With this in mind, we introduce the following assumption that is analogous to standard regularity conditions for GMM.

Assumption GMM1. (a) Θ is compact; (b) $\theta_0 \in \mathrm{int}(\Theta)$ is the unique solution to $G(\theta) = 0$ a.s.; (c) $\Xi_n \xrightarrow{\mathbb{P}} \Xi$ for some $\mathcal{F}$-measurable $k \times k$ matrix Ξ, which is a.s. finite and positive definite.

Assumption GMM1 is mainly concerned with the identification of the true parameter θ_0. The key assumption is the uniqueness of θ_0 as a solution to $G(\theta) = 0$ a.s., which, combined with the positive definiteness of Ξ, ensures that θ_0 is the unique minimizer of $Q(\cdot)$.

To study the asymptotic behavior of $\hat{\theta}_n$, as well as other estimators below, we need more assumptions.

Assumption H1. (a) The processes Y^*, Z and X are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We suppose that X is an Itô semimartingale with the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{\mathbb{R}} \delta(s,z)\, \mu(ds,dz),$$

with b_t locally bounded and σ_t càdlàg. Moreover, $|\delta(\omega, t\wedge\tau_m(\omega), z)| \wedge 1 \leq \Gamma_m(z)$ for all (ω, t, z), where (τ_m) is a localizing sequence of stopping times and for some $r \in (0,1)$, each function Γ_m on $\mathbb{R}$ satisfies $\int_{\mathbb{R}} \Gamma_m(z)^r \lambda(dz) < \infty$.

(b) The process σ_t is also an Itô semimartingale with the form

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

where $\tilde{b}_t$, $\tilde{\sigma}_t$ and $\tilde{\sigma}'_t$ are locally bounded càdlàg adapted, W' a Brownian motion orthogonal to W, $\tilde{\delta}$ is a predictable function satisfying $|\tilde{\delta}\left(\omega, t \wedge \tau_m\left(\omega\right), z\right)| \wedge 1 \leq \tilde{\Gamma}_m\left(z\right)$ for all (ω, t, z), and each function $\tilde{\Gamma}_m$ on $\mathbb{R}$ satisfies $\int_{\mathbb{R}} \tilde{\Gamma}_m\left(z\right)^2 \lambda\left(dz\right) < \infty$.

(c) The processes Y^* and Z are Itô semimartingales satisfying a similar condition as σ_t in part (b).

Assumption H1 is fairly unrestrictive and satisfied by most continuous-time models in finance. We allow the price jumps to have finite or infinite activity, volatility jumps with arbitrary activity level, leverage effect and arbitrary dependence among various components of the price process. The setting here is of course not completely general. For example, we restrict the price jumps to have finite variation, which is a standard condition (see Jacod and Protter (2012)) to ensure that the truncation-based spot variance estimator is robust to jumps in the analysis of second-order asymptotics. The Itô semimartingale model for the stochastic volatility excludes processes such as the fractional Brownian motion.

Assumption H2. For some $p \geq 3$, we have $\varpi \in [(4p-1)/2(4p-r), 1/2)$; recall r in Assumption H1(a). Moreover, for any compact $\mathcal{K}_{\mathcal{Z}} \subseteq \mathcal{Z}$, there exists some constant $C > 0$ such that for all $v \in \mathcal{V}$,

(a) for each $\theta \in \Theta$, $\sup_{z \in \mathcal{K}_{\mathcal{Z}}} \left(\|\partial_v^m d_j\left(z, v; \theta\right)\| + |\partial_v^m f_j\left(z, v; \theta\right)|\right) \leq C\left(1 + |v|^{p-m}\right)$ for $m = 0, 1, 2, 3$;

(b) $\sup_{\theta \in \Theta, z \in \mathcal{K}_{\mathcal{Z}}} \left(\|\partial_\theta^m d_j\left(z, v; \theta\right)\| + \|\partial_\theta^m f_j\left(z, v; \theta\right)\|\right) \leq C\left(1 + |v|^p\right)$, $m = 1, 2$;

(c) $\sup_{\theta \in \Theta, z \in \mathcal{K}_{\mathcal{Z}}} \left(\|\partial_\theta \partial_v d_j\left(z, v; \theta\right)\| + \|\partial_\theta \partial_v^m f_j\left(z, v; \theta\right)\|\right) \leq C\left(1 + |v|^{p-1}\right)$, $m = 1, 2$;

(d) all partial derivatives in (a), (b) and (c) are continuous in (z, v, θ).

Assumption H2 imposes smoothness on $d_j(\cdot)$ and $f_j(\cdot)$. We need $d_j(\cdot)$, $f_j(\cdot)$ and their derivatives to have at most polynomial growth in the spot variance. This condition is needed to tame the effect of the estimation error of the spot variance in the second-stage GMM estimation and easy to verify in practice. For example, if $f_j(z, v; \theta) = \sum_{m=0}^{\bar{m}} c_m(z, \theta) v^m$, $\bar{m} \geq 0$, then $f_j(\cdot)$ verifies the conditions as soon as $c_m(\cdot)$ is twice differentiable in θ with continuous partial derivatives.

Assumption H3. Conditionally on $\mathcal{F}$, the variables $(\varepsilon_t)_{t\in[0,T]}$ are mutually independent with mean zero, where $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{J,t})$. The process $t \mapsto \mathbb{E}\left[\|\varepsilon_t\|^4|\mathcal{F}\right]$ is $\mathcal{F}_t$-adapted and locally bounded. Moreover, $A_t \equiv \mathbb{E}\left[\varepsilon_t\varepsilon_t^\intercal|\mathcal{F}\right]$ is $\mathcal{F}_t$-adapted and càdlàg.

The conditional independence condition on ε_t is somewhat strong in that it rules out autocorrelation of the pricing error. This assumption is akin to the assumption employed in the estimation of volatility and jump functionals that are robust to the microstructure noise; see Jacod et al. (2009), Jacod et al. (2010) and Aït-Sahalia et al. (2012). This said, Assumption H3 does allow for temporal heteroskedasticity in ε_t, unconditional dependence between ε_t and the underlying processes as well as serial dependence in the ε_t process itself through higher moments.

Assumption H4. $k_n^2\Delta_n \to 0$ and $k_n^3\Delta_n \to \infty$.

Assumption H4 imposes undersmoothing ($k_n^2\Delta_n \to 0$); it is well-known that the optimal pointwise estimation of the spot variance demands $k_n \asymp \Delta_n^{-1/2}$ in the basic setting where X is continuous. Undersmoothing is needed here to facilitate a feasible inference procedure; see Jacod and Rosenbaum (2012) for a detailed analysis.

To simplify the exposition, we maintain Assumptions GMM1, H1-H4, as well as GMM2 and GMM3 below throughout the paper without further mention. Under these assumptions, $\hat{\theta}_n$ is a consistent estimator for θ_0, as shown below.

Theorem 1 *(a) We have (2.3) uniformly in $\theta \in \Theta$; (b) $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.*

2.2 Second order bias in $\hat{\theta}_n$

We now consider the second-order properties of $\hat{\theta}_n$. By routine manipulation[1], $\hat{\theta}_n$ admits the following representation with probability approaching one:

$$\hat{\theta}_n - \theta_0 = -(\widehat{H}_n(\hat{\theta}_n)^\intercal \Xi_n \widehat{H}_n(\bar{\theta}_n))^{-1} \widehat{H}_n(\hat{\theta}_n)^\intercal \Xi_n \widehat{G}_n(\theta_0), \tag{2.4}$$

where $\widehat{H}_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_\theta g(Y_{i\Delta_n}, X_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)$ is the $k \times \dim(\theta)$ dimensional Jacobian matrix associated with the sample moment function $\widehat{G}_n(\theta)$ and $\bar{\theta}_n$ is a mean value on the line joining θ_0 and $\hat{\theta}_n$ which may change from element to element in the vector equation. The asymptotic property of $\hat{\theta}_n$ is determined by the asymptotic property of $\widehat{H}_n(\cdot)$ in a neighborhood of θ_0 and the asymptotic property of $\widehat{G}_n(\theta_0)$. To describe the asymptotic behaviors of $\widehat{H}_n(\cdot)$ and $\widehat{G}_n(\theta_0)$, we

[1]See Newey and McFadden (1994) and the proof of Proposition 1 for details.

set

$$
\begin{aligned}
H(\theta) \equiv \int_0^T \partial_\theta g\left(Y_s^*, Z_s, V_s; \theta\right) ds, \quad H \equiv H(\theta_0) \\
B \equiv \int_0^T \partial_v^2 g\left(Y_s^*, Z_s, V_s; \theta_0\right) V_s^2 ds.
\end{aligned} \tag{2.5}
$$

Assumption GMM2. $H^\intercal \Xi H$ is a.s. nonsingular.

Proposition 1 *Let Θ_0 be some neighborhood containing θ_0. We have*

(a) $\sup_{\theta \in \Theta_0} \|\widehat{H}_n(\theta) - H(\theta)\| = o_p(1)$*;*

(b) $k_n \widehat{G}_n(\theta_0) \xrightarrow{\mathbb{P}} B$*;*

(c) $k_n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathbb{P}} -(H' \Xi H)^{-1} H' \Xi B$*.*

Proposition 1 illustrates the standard, as well as the non-standard, aspects of the asymptotic behavior of $\hat{\theta}_n$. Proposition 1(a) is analogous to the convergence of the Jacobian matrix in a standard GMM setting. Although $\widehat{V}_{i\Delta_n}$ is only a noisy proxy of $V_{i\Delta_n}$, the approximation is precise enough so that the approximation error is negligible in the first-order asymptotics of $\widehat{H}_n$. However, in contrast to the standard GMM setting, part (b) shows that the scaled sample moment $k_n \widehat{G}_n(\theta_0)$ does not fulfill a CLT; instead, it converges in probability to a (random) bias term B. This bias arises from the approximation of $V_{i\Delta_n}$ via the spot variance estimator $\widehat{V}_{i\Delta_n}$, coupled with the nonlinearity of $g(y, z, v; \theta_0)$ in v. We note that the function $g(y, z, v; \theta_0)$ is in general nonlinear in v even if $f_j(z, v; \theta_0)$ is linear in v, as the instrument $d_j(z, v; \theta)$ may and generally does depend on v. Part (c) is a direct consequence of parts (a,b). Part (c) illustrates the second-order bias of the GMM estimator $\hat{\theta}_n$. The lack of CLT for $\hat{\theta}_n$ renders standard inference impossible. Importantly, it invalidifies the "naive" approach of treating $\widehat{V}_{i\Delta_n}$ as if it is $V_{i\Delta_n}$ without error. This said, the positive message of Proposition 1(c) is that $\hat{\theta}_n$ is a k_n-consistent estimator for θ_0—an improvement relative to the consistency result in Theorem 1. The k_n-consistency is useful in the discussion of the bias correction for $\hat{\theta}_n$.

For concreteness, we use a simple linear regression model as a running example throughout the paper.

Linear Regression Example. Let $T = 1$, $\theta = (\alpha, \beta)$ and $f(v; \theta) = \alpha + \beta v$. In an ordinary least square (OLS) settting, the intruments are $(1, v)$, so we have $g(y, v; \theta) = (\alpha + \beta v - y, v(\alpha + \beta v - y))$. Let $IV = \int_0^1 V_s ds$ and $IQ = \int_0^1 V_s^2 ds$ be respectively the integrated variance and quarticity.

Proposition 1 shows

$$k_n \begin{pmatrix} \hat{\alpha}_n - \alpha_0 \\ \hat{\beta}_n - \beta_0 \end{pmatrix} \xrightarrow{\mathbb{P}} -H^{-1}B = \frac{2\beta_0}{1 - IV^2/IQ} \begin{pmatrix} IV \\ -1 \end{pmatrix}.$$
$$\text{where} \quad H = \begin{pmatrix} 1 & IV \\ IV & IQ \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 2\beta_0 IQ \end{pmatrix}.$$

By the Cauchy-Schwarz inequality, $IQ \geq IV^2$ and the equality holds if and only if V_t is constant for $t \in [0, T]$. Therefore, the relative bias (normalized by β_0) in the slope estimation is approximately $-2k_n^{-1}(1 - IV^2/IQ)^{-1} < 0$. Hence, the bias is towards zero and the effect is large when IQ is close to IV^2, i.e., V_t has low variation over $[0, T]$. □

2.3 Bias correction

In this subsection, we propose a simple bias-correction for $\hat{\theta}_n$ and derive the limit distribution of the bias-corrected estimator. By computing the derivatives in (2.5) explicitly and using the fact that $Y_t^* = f(Z_t, V_t; \theta_0)$, it is easy to see that $B = \int_0^T \gamma(Z_s, V_s; \theta_0)\, ds$, where $\gamma(z, v; \theta) = (\gamma_1(z, v; \theta), \ldots, \gamma_J(z, v; \theta))$ and for $1 \leq j \leq J$,

$$\gamma_j(z, v; \theta) = \left(2\partial_v d_j(z, v; \theta)\, \partial_v f_j(z, v; \theta) + d_j(z, v; \theta)\, \partial_v^2 f_j(z, v; \theta)\right) v^2.$$

We set

$$\widehat{B}_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \gamma\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right),$$

and consider a general class of bias-corrected estimators, henceforth the *two-step estimators*, as follows:

$$\theta_n^*(\widetilde{H}_n, \tilde{\theta}_n) \equiv \hat{\theta}_n + \frac{1}{k_n}\left(\widetilde{H}_n^\intercal \Xi_n \widetilde{H}_n\right)^{-1} \widetilde{H}_n^\intercal \Xi_n \widehat{B}_n(\tilde{\theta}_n), \tag{2.6}$$

where $\widetilde{H}_n$ and $\tilde{\theta}_n$ are preliminary estimators which are "close to" $\widehat{H}_n(\hat{\theta}_n)$ and θ_0, respectively.

Our main theorem in this subsection is on the asymptotic distribution of $\theta_n^*(\widetilde{H}_n, \tilde{\theta}_n)$. To describe the asymptotic variance, we consider a $k \times k$ matrix $S_1(\theta)$ with its (j, l) block, $1 \leq j, l \leq J$, given by

$$\int_0^T d_j(Z_s, V_s; \theta)\, d_l(Z_s, V_s; \theta)^\intercal A_{jl,s} ds.$$

We also set for $1 \leq j \leq J$, $\phi_j(z, v; \theta) \equiv d_j(z, v; \theta)\, \partial_v f_j(z, v; \theta)$ and $\phi(z, v; \theta) = (\phi_1(z, v; \theta), \ldots, \phi_J(z, v; \theta))$ and consider the $k \times k$ matrix

$$S_2(\theta) \equiv 2\int_0^T \phi(Z_s, V_s; \theta)\, \phi(Z_s, V_s; \theta)^\intercal V_s^2 ds.$$

Finally, let $S(\theta) \equiv S_1(\theta) + S_2(\theta)$ and $S \equiv S(\theta_0)$.

Theorem 2 *Suppose that* $\widetilde{H}_n = \widehat{H}_n(\hat{\theta}_n) + O_p\left(k_n^{-1}\right)$ *and* $\tilde{\theta}_n = \theta_0 + O_p\left(k_n^{-1}\right)$. *Let* $\theta_n^* = \theta_n^*(\widetilde{H}_n, \tilde{\theta}_n)$ *as in (2.6). We have*

$$\Delta_n^{-1/2}(\theta_n^* - \theta_0) = M\xi_n + o_p(1), \tag{2.7}$$

where $M \equiv -\left(H^\intercal \Xi H\right)^{-1} H^\intercal \Xi$ *and the sequence of variables* $(\xi_n)_{n\geq 1}$*, which does not depend on* $\widetilde{H}_n$ *and* $\tilde{\theta}_n$*, converges stably in law to a variable defined on an extension of the space* $(\Omega, \mathcal{F}, \mathbb{P})$*, which conditionally on* $\mathcal{F}$*, is centered Gaussian with variance-covariance matrix* S*.*

Theorem 2 shows that the bias-corrected estimator θ_n^* has an asymptotic representation $\Delta_n^{-1/2}(\theta_n^* - \theta_0) = M\xi_n + o_p(1)$. In particular, $\Delta_n^{-1/2}(\theta_n^* - \theta_0)$ converges stably in law towards a mixture centered Gaussian distribution with $\mathcal{F}$-conditional variance $\Sigma = MSM^\intercal$. Moreover, the theorem shows that the bias-corrected estimators with various choices of $\widetilde{H}_n$ and $\tilde{\theta}_n$ are second-order asymptotically equivalent, in that they share the same asymptotic representation (2.7).

The asympototic variance Σ has a standard format for GMM estimators:

$$\Sigma = \left(H^\intercal \Xi H\right)^{-1} H^\intercal \Xi S \Xi H \left(H^\intercal \Xi H\right)^{-1}. \tag{2.8}$$

As is well known, the optimal choice of the weighting matrix is $\Xi = S^{-1}$, yielding $\Sigma = \left(H^\intercal S^{-1} H\right)^{-1}$. We note that the matrix S contains two components: S_1 captures the sampling variability arising from the random disturbance ε_t and S_2 accounts for the sampling error in the estimation of spot variances. It is not surprising that the first-stage estimation of the spot variance affects the asymptotic variance in the second stage. This said, the problem here is quite different than a two-stage estimation procedure commonly seen in a GMM setting, because the first-stage estimation is nonparametric in nature, but not kernel- or sieve-based, and is constructed from (possibly discontinuous) semimartingales instead of independent or weakly dependent data as is typical in cross-sectional or macro time-series settings.

Linear Regression Example—Continued. In the OLS setting, S_1 and S_2 are given as follows:

$$S_1 = \int_0^1 \begin{pmatrix} A_s & V_s A_s \\ V_s A_s & V_s^2 A_s \end{pmatrix} ds, \quad S_2 = 2\beta_0^2 \int_0^1 \begin{pmatrix} V_s^2 & V_s^3 \\ V_s^3 & V_s^4 \end{pmatrix} ds;$$

recall that $A_s = \mathbb{E}\left[\varepsilon_s^2 | \mathcal{F}\right]$ is the conditional variance of the random error. □

We now turn to feasible choices of $\widetilde{H}_n$ and $\tilde{\theta}_n$ that verify the condition in Theorem 2. By Proposition 1(c), the raw estimator $\hat{\theta}_n$ is k_n-consistent and thus is a valid candidate for $\tilde{\theta}_n$. More generally, any bias-corrected estimator θ_n^* is $\Delta_n^{-1/2}$-, hence also k_n-consistent, and can be used as a valid choice for $\tilde{\theta}_n$. An obvious choice for $\widetilde{H}_n$ is to set $\widetilde{H}_n = \widehat{H}_n(\hat{\theta}_n)$, which trivially verifies the condition in Theorem 2. We also propose an alternative choice, which is easier to compute in applications. To this end, we set $\psi_j(z, v; \theta) = d_j(z, v; \theta)\, \partial_\theta f_j(z, v; \theta)^\intercal$, $1 \leq j \leq J$, $\psi(z, v; \theta) = (\psi_1(z, v; \theta), \ldots, \psi_J(z, v; \theta))$; ψ_j is $k_j \times \dim(\theta)$ and ψ is $k \times \dim(\theta)$. By direct calculation, $H = \int_0^T \psi(Z_s, V_s; \theta_0) ds$. We hence consider

$$\widehat{H}_n^*(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \psi\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right), \quad \theta \in \Theta.$$

Lemma 1 *Let θ_n^H be a preliminary estimator satisfying $\theta_n^H = \theta_0 + O_p\left(k_n^{-1}\right)$. Then (a) $\widehat{H}_n(\theta_n^H) = \widehat{H}_n(\hat{\theta}_n) + O_p(k_n^{-1})$; (b) $\widehat{H}_n^*(\theta_n^H) = \widehat{H}_n(\hat{\theta}_n) + O_p(k_n^{-1})$.*

Comment. Lemma 1 shows that $\widehat{H}_n(\theta_n^H)$ and $\widehat{H}_n^*(\theta_n^H)$ are both valid candidates for H_n^* that verify the condition in Theorem 2, provided that θ_n^H is a k_n-consistent estimator for θ_0. Feasible choices of θ_n^H include $\hat{\theta}_n$ or any bias-corrected estimator θ_n^*.

Linear Regression Example—Continued. Let the raw estimator be $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$. The instrument in the OLS setting is $(1, v)$. Hence,

$$\widehat{H}_n^*(\theta) = \begin{pmatrix} \iota_n & IV_n \\ IV_n & IQ_n \end{pmatrix}, \quad \widehat{B}_n(\theta) = \begin{pmatrix} 0 \\ 2\beta IQ_n \end{pmatrix}$$

where $\iota_n = \Delta_n([1/\Delta_n] - k_n + 1)$, $IV_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \widehat{V}_{i\Delta_n}$ and $IQ_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \widehat{V}_{i\Delta_n}^2$. In the simple OLS setting, $\widehat{H}_n^*(\theta)$ does not depend on θ. A bias-corrected estimator $\theta_n^* = (\alpha_n^*, \beta_n^*)$ can be constructed as

$$\begin{aligned} \theta_n^* &= \hat{\theta}_n + \frac{1}{k_n}\left(\widehat{H}_n^*\right)^{-1} \widehat{B}_n\left(\hat{\theta}_n\right) \\ &= \hat{\theta}_n + \frac{1}{k_n} \frac{2\hat{\beta}_n}{\iota_n - IV_n^2/IQ_n} \begin{pmatrix} -IV_n \\ 1 \end{pmatrix}. \end{aligned}$$

One can iterate the bias-correction by using $\widehat{B}_n(\theta_n^*)$ and get

$$\theta_n^{**} = \hat{\theta}_n + \frac{1}{k_n} \frac{2\beta_n^*}{\iota_n - IV_n^2/IQ_n} \begin{pmatrix} -IV_n \\ 1 \end{pmatrix}.$$

2.4 Estimation of asymptotic variance and efficient GMM

For the purpose of making inference on the basis of Theorem 2, we need a consistent estimator for the asymptotic variance Σ. We start with constructing an estimator for S. Let $\hat{\varepsilon}_{j,i\Delta_n}(\theta) = Y_{j,i\Delta_n} - f_j(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)$. We define $\widehat{S}_{1,n}(\theta)$ to be a $k \times k$ matrix with its (j,l) block given by

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} d_j\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right) d_l\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right)^{\intercal} \hat{\varepsilon}_{j,i\Delta_n}(\theta) \hat{\varepsilon}_{l,i\Delta_n}(\theta)$$

and set

$$\widehat{S}_{2,n}(\theta) \equiv 2\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \phi\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right) \phi\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta\right)^{\intercal} \widehat{V}_{i\Delta_n}^2.$$

We then set $\widehat{S}_n(\theta) \equiv \widehat{S}_{1,n}(\theta) + \widehat{S}_{2,n}(\theta)$. Finally, for some preliminary estimator $\widetilde{H}_n$ for H, we set $\widetilde{M}_n \equiv (\widetilde{H}_n^{\intercal}\Xi_n\widetilde{H}_n)^{-1}\widetilde{H}_n^{\intercal}\Xi_n$ and $\widehat{\Sigma}_n(\widetilde{H}_n, \tilde{\theta}_n) \equiv \widetilde{M}_n\widehat{S}_n(\tilde{\theta}_n)\widetilde{M}_n^{\intercal}$,

Theorem 3 *Suppose* $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$ *and* $\widetilde{H}_n \xrightarrow{\mathbb{P}} H$. *Then* $\widehat{S}_n(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} S$ *and* $\widehat{\Sigma}_n(\widetilde{H}_n, \tilde{\theta}_n) \xrightarrow{\mathbb{P}} \Sigma$.

COMMENTS. (i) To construct a consistent estimator for Σ, one can take $\tilde{\theta}_n$ to be the raw estimator $\hat{\theta}_n$ and $\widetilde{H}_n$ to be $\widehat{H}_n(\hat{\theta}_n)$ or $\widehat{H}_n^*(\hat{\theta}_n)$. More generally, $\hat{\theta}_n$ can be replaced by any bias-corrected estimator, such as θ_n^*.

(ii) By the properties of stable convergence in law[2], Theorems 2 and 3 imply that $\Delta_n^{-1/2}\widehat{\Sigma}_n(\widetilde{H}_n, \tilde{\theta}_n)^{-1/2}(\theta_n^* - \theta_0)$ converges in distribution to $\mathcal{N}(0, I_{\dim(\theta)})$. Confidence sets for θ_0, or its subvector, can be constructed in the usual way.

An efficient GMM procedure, in the sense of minimizing Σ by optimally choosing Ξ, can be implemented by setting $\Xi_n = \widehat{S}_n(\tilde{\theta}_n)^{-1}$ where the preliminary estimator $\tilde{\theta}_n$ verifies $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$. The two-step estimator θ_n^* (see Section 2.3) associated with the optimal weighting matrix is referred to as the *efficient two-step estimator*.

2.5 One-step estimators

We now introduce an alternative estimator based on a bias-corrected objection function. We set $Q_n^\star(\theta) = G_n^\star(\theta)^{\intercal}\Xi_n G_n^\star(\theta)$, where the bias-corrected moment function is $G_n^\star(\theta) \equiv \widehat{G}_n(\theta) - k_n^{-1}\widehat{B}_n(\theta)$. The bias correction is motivated by Proposition 1. The *one-step estimator* is given by

$$\theta_n^\star = \arg\min_{\theta\in\Theta} Q_n^\star(\theta). \tag{2.9}$$

[2] For random variables ξ_n and ζ_n, if ξ_n converges stably in law to ξ and $\zeta_n \xrightarrow{\mathbb{P}} \zeta$, then (ξ_n, ζ_n) converge stably in law to (ξ, ζ). See e.g. (2.2.5) in Jacod and Protter (2012).

The following theorem shows that $\theta_n^\star$ has the same asymptotic representation as θ_n^*. We call $\theta_n^\star$ the efficient one-step estimator if it is associated with the optimal weighting matrix $\Xi = S^{-1}$.

Theorem 4 *Consider the same setting as in Theorem 2. Let $\theta_n^\star$ be given by (2.9) with weighting matrix $\Xi_n^\star$ such that $\Xi_n^\star \xrightarrow{\mathbb{P}} \Xi$. We have $\Delta_n^{-1/2}(\theta_n^\star - \theta_0) = M\xi_n + o_p(1)$.*

Linear Regression Example—Continued. Let $\overline{Y}_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} Y_{i\Delta_n}$ and $VY_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \widehat{V}_{i\Delta_n} Y_{i\Delta_n}$. We have

$$\begin{aligned}
\widehat{G}_n(\theta) &= \begin{pmatrix} \iota_n & IV_n \\ IV_n & IQ_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \overline{Y}_n \\ VY_n \end{pmatrix} \\
G_n^\star(\theta) &= \begin{pmatrix} \iota_n & IV_n \\ IV_n & \left(1 - \frac{2}{k_n}\right) IQ_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \overline{Y}_n \\ VY_n \end{pmatrix}.
\end{aligned}$$

The raw estimator and the one-step estimator are respetively given by

$$\begin{aligned}
\hat{\theta}_n &= \begin{pmatrix} \iota_n & IV_n \\ IV_n & IQ_n \end{pmatrix}^{-1} \begin{pmatrix} \overline{Y}_n \\ VY_n \end{pmatrix} \\
\theta_n^\star &= \begin{pmatrix} \iota_n & IV_n \\ IV_n & \left(1 - \frac{2}{k_n}\right) IQ_n \end{pmatrix}^{-1} \begin{pmatrix} \overline{Y}_n \\ VY_n \end{pmatrix}.
\end{aligned}$$

The one-step estimator can be modified to a *continuous updating estimator* (Hansen et al. (1996)), which achieves the bias correction and the optimal choice of weighting matrix within the estimation step. We consider $Q_n^c(\theta) = G_n^\star(\theta)^\intercal \widehat{S}_n(\theta)^{-1} G_n^\star(\theta)$ and the continuous updating estimator is given by

$$\theta_n^c \equiv \arg\min_{\theta\in\Theta} Q_n^c(\theta).$$

We need an additional assumption.

Assumption GMM3. $\inf_{\theta\in\Theta} \lambda_{\min}(S(\theta)) > 0$ a.s., where $\lambda_{\min}(\cdot)$ is the smallest eigenvalue function.

Assumption GMM3 requires that the eigenvalues of $S(\cdot)$ are positive and uniformly bounded away from zero; this ensures (i) $S(\theta)$ is positive definite over the entire parameter space, so that θ_0 is the unique minimizer of the limit objective function; (ii) $\widehat{S}_n(\theta)^{-1} \xrightarrow{\mathbb{P}} S(\theta)^{-1}$ uniformly in θ.

Theorem 5 *We have* $\Delta_n^{-1/2}(\theta_n^c - \theta_0) = -(H^\intercal S^{-1}H)^{-1}H^\intercal S^{-1}\xi_n + o_p(1)$, *where* ξ_n *is the same as in Theorem 2.*

Theorem 4 shows that the one-step estimator $\theta_n^\star$ has the same asymptotic representation as θ_n^*; these estimators hence have the same asymptotic distribution. Theorem 5 shows that the continuous updating estimator is asymptotically equivalent to the efficient one-step and two-step estimators on the second order. Conducting the bias-correction within the estimation step is a desirable feature for the one-step estimator and the continuous updating estimator. As shown below, the bias-corrected moment condition $G_n^\star$ and objective functions $Q_n^\star$ and Q_n^c can be used for constructing tests and confidence sets that are robust to weak identification.

This said, the "automatic" bias-correction is achieved at the cost of minimizing a more complicated objective function. Indeed, in order to compute $G_n^\star(\cdot)$, one needs to evaluate $\partial_v d_j(\cdot)$, $\partial_v f_j(\cdot)$ and $\partial_v^2 f(\cdot)$ in every trial for each observation in the optimization algorithm. Evaluating derivatives is not a problem when $f_j(\cdot)$ and $d_j(\cdot)$ can be written in closed form. However, in many empirical option pricing applications, the pointwise evaluation of $f_j(\cdot)$ often involves complicated numerical procedures, such as Fourier transform, numerical solution to ordinary differential equations, and sometimes a large number of Monte Carlo simulations. All such numerical complications render precise evaluations of the derivatives of $f_j(\cdot)$ and $d_j(\cdot)$ time consuming and maybe unpractical. When numerical complexity is of concern, we recommend the use of the two-step estimator θ_n^*, which is without loss of statistical efficiency relative to its one-step counterpart.

3 Inference

3.1 Overidentification test

The bias-corrected objective function $Q_n^\star$ can be used to construct Hansen's (1982) J statistic testing the overidentifying conditions.

Theorem 6 *Suppose* $\Xi = S^{-1}$. *Let* θ_n^* *and* $\theta_n^\star$ *be given as in Theorems 2 and 4. Then* $\Delta_n^{-1}Q_n^\star(\theta_n^*) \xrightarrow{d} \chi^2_{k-\dim(\theta)}$ *and the same convergence holds for* $\Delta_n^{-1}Q_n^\star(\theta_n^\star)$ *and* $\Delta_n^{-1}Q_n^c(\theta_n^c)$.

Comments. (i) The J-statistic can be constructed based on either the efficient one-step estimator or the efficient two-step estimator. More generally, any estimator ϑ_n satisfying $\Delta_n^{-1/2}(\theta_n^\star - \theta_0) = M\xi_n + o_p(1)$ can be used for the same purpose.

(ii) The J statistic is evaluated with the bias-corrected objective function $Q_n^\star$. The objective function $\widehat{Q}_n$ can not be used for such purpose. Indeed, it can be shown that $k_n^2\widehat{Q}_n(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} B^\intercal \Xi B$ for any $\tilde{\theta}_n$ satisfying $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

3.2 Parameter stability

We now consider a multi-period problem. The time span of period τ, $\tau = 1, \ldots, \bar{\tau}$, is $[(\tau - 1)\,T, \tau T)$ with $T > 0$. In each period τ, we assume that the true parameter $\theta_0(\tau)$ is constant but allow $\theta_0(\tau)$ to vary in τ. We are interested in constructing the uniform confidence band for the parameter path $\tau \mapsto \kappa^\intercal \theta_0(\tau)$, where $\kappa = (\kappa_j)$ is a $\dim(\theta) \times 1$ constant vector. For example, setting $\kappa_j = 1_{\{j=l\}}$ gives the path of the lth component of θ_0. We denote by $\theta_n^*(\tau)$ the two-step estimator based on period-τ data (see Theorem 2) and $\widehat{\Sigma}_n(\tau)$ be a consistent estimator for $\Sigma(\tau)$, the latter being the asymptotic variance associated with $\theta_n^*(\tau)$. For $\alpha \in (0,1)$, we set $z_{\bar{\tau},\alpha}$ to be the $1-\alpha/2$ quantile of the variable $\max_{1\le\tau\le\bar{\tau}} |\mathcal{N}_\tau|$, where $\mathcal{N}_\tau$, $1 \le \tau \le \bar{\tau}$, are independent standard normal variables. A nominal level α uniform confidence band for $\kappa^\intercal\theta_0(\tau)$ is given by the following sequence of random intervals: for $1 \le \tau \le \bar{\tau}$,

$$CI_n^*(\tau;\kappa,\alpha) \equiv \left[\kappa^\intercal\theta_n^*(\tau) - z_{\bar{\tau},\alpha}\sqrt{\kappa^\intercal\widehat{\Sigma}_n(\tau)\,\kappa}, \kappa^\intercal\theta_n^*(\tau) + z_{\bar{\tau},\alpha}\sqrt{\kappa^\intercal\widehat{\Sigma}_n(\tau)\,\kappa}\right].$$

It is easy to see that similar constructions can be based on the one-step estimator and the continuous updating estimator. We omit the details for brevity. The asymptotic property of above confidence band is formally given by the theorem below.

Theorem 7 *The variables $\widehat{\Sigma}_n(\tau)^{-1/2}(\theta_n^*(\tau) - \theta_0(\tau))$, $1 \le \tau \le \bar{\tau}$, converge in distribution to a $\bar{\tau}$-dimensional standard normal distribution. Moreover, for any $\alpha \in (0,1)$ and $\dim(\theta) \times 1$ vector $\kappa \neq 0$,*

$$\mathbb{P}\left(\kappa^\intercal\theta_0(\tau) \in CI_n^*(\tau;\kappa,\alpha) \text{ for all } 1 \le \tau \le \bar{\tau}\right) \to \alpha.$$

The uniform confidence band is a natural tool for visualizing the temporal stability/instability of the one-dimensional parameter $\kappa^\intercal\theta_0(\tau)$. In practice, we are also interested in testing for the joint stability of the vector $\theta_0(\tau)$. Such a test can be formalized as the following hypotheses:

$$\begin{cases} H_0 : \theta_0(1) = \cdots = \theta_0(\bar{\tau}) \\ H_a : \theta_0(\tau) \neq \theta_0(\tau'), \text{ some } 1 \le \tau, \tau' \le \bar{\tau}. \end{cases} \tag{3.1}$$

To construct formal tests, we set $\Delta\theta_n^*(\tau) = \Delta_n^{-1/2}(\theta_n^*(\tau) - \theta_n^*(\tau+1))$, $1 \le \tau \le \bar{\tau} - 1$, and

$$\underset{(\bar{\tau}-1)\times\bar{\tau}}{R} = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix},$$

$$\widehat{\mathbf{\Sigma}}_n = \begin{pmatrix} \widehat{\Sigma}_n(1) & & \\ & \ddots & \\ & & \widehat{\Sigma}_n(\bar{\tau}) \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \Sigma(1) & & \\ & \ddots & \\ & & \Sigma(\bar{\tau}) \end{pmatrix}.$$

We consider test statistics of the following form:

$$PS_n = \mathcal{S}\left(\Delta\theta_n^*(1), \ldots, \Delta\theta_n^*(\bar{\tau}-1); \widehat{\mathbf{\Sigma}}_n\right).$$

where the test function $\mathcal{S} : \mathbb{R}^{(\bar{\tau}-1)\dim(\theta)} \times \mathbb{R}^{\dim(\theta)\times\dim(\theta)\times\bar{\tau}} \mapsto \mathbb{R}_+$ satisfies the following properties: (i) $\mathcal{S}(\cdot)$ is continuous on $\mathbb{R}^{(\bar{\tau}-1)\dim(\theta)} \times \mathcal{M}_+^{\bar{\tau}}$, where $\mathcal{M}_+$ is the space of positive definite $\dim(\theta) \times \dim(\theta)$ matrices; (ii) $\mathcal{S}(x_1, \ldots, x_{\bar{\tau}-1}; \mathbf{\Sigma}) = 0$ if and only if $x_\tau = 0$ for all $1 \le \tau \le \bar{\tau}-1$; (iii) $\mathcal{S}(x_1, \ldots, x_{\bar{\tau}-1}; \mathbf{\Sigma}) \to \infty$ whenever some $\|x_\tau\| \to \infty$ for some $1 \le \tau \le \bar{\tau}-1$. We list two examples for concreteness:

$$\mathcal{S}\left(\Delta\theta_n^*(1), \ldots, \Delta\theta_n^*(\bar{\tau}-1); \widehat{\mathbf{\Sigma}}_n\right) = \begin{cases} \max_{\tau\in\mathcal{T}} \left\| \left(\widehat{D}_n(\tau) + \widehat{D}_n(\tau+1)\right)^{-1/2} \Delta\theta_n^*(\tau) \right\| \\ \max_{\tau\in\mathcal{T}} \Delta\theta_n^*(\tau)^\intercal \left(\widehat{\Sigma}_n(\tau) + \widehat{\Sigma}_n(\tau+1)\right)^{-1} \Delta\theta_n^*(\tau) \end{cases},$$

where $\mathcal{T} \subseteq \{1, \ldots, \bar{\tau}-1\}$ is specified by the user and $\widehat{D}_n(\tau)$ collects the diagonal elements of $\widehat{\Sigma}_n(\tau)$; note that $\widehat{\Sigma}_n(\tau) + \widehat{\Sigma}_n(\tau+1)$ estimates the asymptotic variance associated with $\Delta\theta_n^*(\tau)$.

We now discuss the asymptotic property of PS_n, followed by implementation details of the test.

Corollary 1 *Consider the hypotheses in (3.1). We have the following.*

(a) Under H_0, PS_n converges stably in law to $\mathcal{S}(\tilde{\xi}; \mathbf{\Sigma})$, for some $(\bar{\tau}-1)\dim(\theta)$-dimensional random vector $\tilde{\xi}$ defined on an extension of the space $(\Omega, \mathcal{F}, \mathbb{P})$, which conditionally on $\mathcal{F}$, is centered Gaussian with variance-covariance matrix $R^\intercal \mathbf{\Sigma} R$.

(b) Let $\tilde{\zeta}$ be a $(\bar{\tau}-1)\dim(\theta)$-dimensional standard normal random variable independent of $\mathcal{F}$. If the $\mathcal{F}$-condtional distribution of $\mathcal{S}(\tilde{\xi}; \mathbf{\Sigma})$ is continuous and strictly increasing, then for any $\alpha \in (0,1)$, the $1-\alpha$ $\mathcal{F}$-conditional quantile of $\mathcal{S}((R^\intercal \widehat{\mathbf{\Sigma}}_n R)^{1/2}\tilde{\zeta}; \widehat{\mathbf{\Sigma}}_n)$ converges in probability to the same conditional quantile of $\mathcal{S}(\tilde{\xi}; \mathbf{\Sigma})$.

(c) Under H_a, PS_n diverges to $+\infty$ in probability.

Corollary 1(a) describes the limit distribution of PS_n under the null hypothesis. A nominal level α test for parameter stability can be carried out by comparing PS_n with a consistent estimator of the $1-\alpha$ $\mathcal{F}$-conditional quantile of the limit distribution. Since the limit null distribution is nonstandard, there is no analytical expression for the quantile in general. Corollary 1(b) shows that the $1-\alpha$ $\mathcal{F}$-conditional quantile of $\mathcal{S}((R^{\intercal}\widehat{\mathbf{\Sigma}}_n R)^{1/2}\tilde{\zeta};\widehat{\mathbf{\Sigma}}_n)$ is a valid estimator for the quantile of the limit distribution; this estimator can be easily computed via simulation by drawing a large number of Monte Carlo realizations of $\tilde{\zeta}$ with $\widehat{\mathbf{\Sigma}}_n$ given. It is easy to see that the critical values form a tight sequence. Hence, part (c) of the corollary implies that the aforementioned test has asymptotic power one under H_a.

3.3 Inference under weak identification

In this subsection, we consider confidence sets for θ_0 which are robust to weak identification. In particular, we show that the methods of Stock and Wright (2000) and Andrews and Soares (2010) can be easily adapted to the current setting. We also consider confidence sets for a subvector $\theta_0^{(1)}$ where we partition θ_0 as $(\theta_0^{(1)},\theta_0^{(2)})$. Here, $\theta_0^{(1)}$ is a component subject to weak identification and $\theta_0^{(2)}$ is strongly identified. For simplicity, we suppose that Θ has a corresponding partition $\Theta_1\times\Theta_2$. The concentrated continuous updating objective function is given by $Q_n^c(\theta^{(1)},\hat{\theta}_n^{(2)}(\theta^{(1)}))$, where

$$\hat{\theta}_n^{(2)}\left(\theta^{(1)}\right)=\arg\min_{\theta^{(2)}\in\Theta_2}Q_n^c\left(\theta^{(1)},\theta^{(2)}\right).$$

The following corollary of Theorem 6 gives the asymptotic distribution of the continuous updating objective function, as well as the concentrated version, evaluated at the true value.

Corollary 2 *We have* $Q_n^c(\theta_0)\xrightarrow{d}\chi_k^2$ *and* $Q_n^c(\theta_0^{(1)},\hat{\theta}_n^{(2)}(\theta_0^{(1)}))\xrightarrow{d}\chi^2_{k-\dim(\Theta_2)}$.

Similarly as in Stock and Wright (2000), we construct Andersen-Rubin type confidence sets by inverting tests, which by construction are robust to weak identification. For $\alpha\in(0,1)$, we set

$$\begin{aligned}CS_{n,1-\alpha}^{SW}&=\left\{\theta\in\Theta:Q_n^c\left(\theta\right)\le\chi^2_{k,1-\alpha}\right\}\\CS_{n,1-\alpha}^{SWSub}&=\left\{\theta^{(1)}\in\Theta_1:Q_n^c\left(\theta^{(1)}\right)\le\chi^2_{k-\dim(\Theta_2),1-\alpha}\right\},\end{aligned}$$

where for any $q\ge1$, $\chi^2_{q,1-\alpha}$ is the $1-\alpha$ quantile of a χ^2_q distribution. Corollary 2 implies that these confidence sets have valid asymptotic coverage:

$$\mathbb{P}\left(\theta_0\in CS_{n,1-\alpha}^{SW}\right)\to\alpha,\quad\mathbb{P}\left(\theta_0^{(1)}\in CS_{n,1-\alpha}^{SWSub}\right)\to\alpha.$$

The method of Andrews and Soares (2010) can also be adapted to the current framework. We discuss one example for concreteness. Let $\widehat{DS}_n\left(\theta\right)$ be the diagonal submatrix of $\widehat{S}_n\left(\theta\right)$. Consider the test statistic:

$$MaxS_n(\theta) = \max_{1 \le j \le \dim(\theta)} \left| \frac{\Delta_n^{-1/2} G^{\star}_{j,n}(\theta)}{\sqrt{\widehat{DS}_{j,n}(\theta)}} \right|.$$

The asymptotic property of $MaxS_n(\theta_0)$ is given below.

Proposition 2 *The variables $MaxS_n(\theta_0)$ converges stably in law to $\max_{1 \le j \le \dim(\theta)} |\tilde{\xi}_j|$, where the $\dim(\theta) \times 1$ variable $\tilde{\xi}$ is defined on an extension of the space $(\Omega, \mathcal{F}, \mathbb{P})$, which conditionally on $\mathcal{F}$, is centered Gaussion with variance-covariance matrix being the correlation matrix induced by S (see Theorem 2).*

A nominal level $1-\alpha$ confidence set can be constructed by inverting tests for the null hypothesis $H_0 : \theta_0 = \theta$. The critical value of the test can be computed as follows. First, estimate the asymptotic variance-covariance matrix of $\tilde{\xi}$ in Proposition 2 by $\Sigma_{\tilde{\xi},n} = \widehat{DS}_n(\theta)^{-1/2} \widehat{S}_n(\theta) \widehat{DS}_n(\theta)^{-1/2}$. Second, draw a large number of Monte Carlo realizations of a random vector $\tilde{\zeta}$ with variance-covariance $\Sigma_{\tilde{\xi},n}$. Third, compute the Monte Carlo $1-\alpha$ quantile of $\max_{1 \le j \le \dim(\theta)} |\tilde{\zeta}|$, denoted by $cv^{AS}_{n,1-\alpha}(\theta)$. By continuous mapping, $cv^{AS}_{n,1-\alpha}(\theta_0)$ consistently estimates the $1-\alpha$ $\mathcal{F}$-conditional quantile of the limit distribution of $MaxS_n(\theta_0)$. The confidence set can then be constructed as

$$CS^{AS}_{n,1-\alpha} = \left\{ \theta \in \Theta : MaxS_n(\theta) \le cv^{AS}_{n,1-\alpha}(\theta) \right\},$$

which has valid asymptotic coverage by Proposition 2.

4 Simulation results

In this section, we examine the asymptotic theory above in two simulation settings that mimic the setup in our empirical application.

4.1 An affine model of the VIX

In the first simulation setting, the logarithm price X_t and the spot variance V_t are generated according to the following stochastic differential equations:

$$\left\{ \begin{array}{rcl} dX_t & = & (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t} dW_t + J_X dN_t \\ dV_t & = & \kappa(\xi - V_t)dt + \sigma\sqrt{V_t} dB_t + J_V dN_t \end{array} \right. \tag{4.1}$$

where W_t and B_t are standard Brownian motions with $\mathbb{E}(dW_t dB_t) = \rho dt$, N_t is a Poisson process with state-dependent intensity $\lambda_t = \lambda_0 + \lambda_1 V_t$, and J_X and J_V are random jump sizes distributed as $J_X \sim N(\mu_X, \sigma_X^2)$ and $J_V \sim exp(\beta_V)$. We set $\kappa = 5$, $\sigma = 0.75$, $\rho = -0.8$, $\xi = 0.06$, $\mu_X = -0.02$, $\sigma_X = 0.05$, $\beta_V = 0.05$, $\lambda_0 = 30$, $\lambda_1 = 60$, $\mu_0 = 0.05$, and $\mu_1 = 0.5$. If the risk-neutral dynamics of (X_t, V_t) also follow (4.1), it can be shown that (see Proposition 3 below) the squared VIX can be written as an affine function of the spot variance, i.e.,

$$Y_t \equiv (VIX_t/100)^2 = a + bV_t + \varepsilon_t. \tag{4.2}$$

We hence simulate Y_t according to this model, henceforth the affine VIX model, with $a = 0.056$ and $b = 0.631$, and $\varepsilon_t \sim IID\mathcal{N}(0, 0.03^2)$; the parameters are calibrated to the dataset employed in Section 5. The parameter of interest here is $\theta = (a, b)$.

Throughout the simulations, we fix $T = 63$ days. We consider two sampling frequencies: $\Delta = 15$ or 60 seconds, where the former setting examines the validity of the asymptotic theory while the latter illustrates potential small-sample distortions. The window size k_n in the spot variance estimation is taken to be $[\bar{\kappa}^{2/5}(\Delta)^{-2/5}]$, with $\bar{\kappa} = 0.5$, 1, and 2. There are 2,000 Monte Carlo trials.

Figure 4.1 presents Monte Carlo distributions of the raw and the two-step (bias-corrected) estimators, non-studentized (left) and studentized (right). Although the raw estimator does not admit a CLT, we studentize the raw estimator with the "naive" standard error which is computed according to standard regression theory treating V_t as if it were estimated without error. In line with the asymptotic theory, the two-step estimator effectively corrects the bias in estimation, and the standardized estimates form a finite-sample distribution well approximated by the standard normal density. These findings are further confirmed by Table 1, which compares the median bias and the median absolute error of the raw and the two-step estimates, and Table 2, which compares the finite-sample rejection probabilities for 5% and 10% nominal level tests. Although the two-step estimator exhibits mild over-rejection, it provides significant improvement over its "naive" counterpart.

4.2 Alternative exponential-affine specification

We now consider a nonlinear specification for VIX. Following Huang and Tauchen (2005) and Barndorff-Nielsen et al. (2008), we simulate

$$\begin{aligned} dX_t &= \mu dt + \sqrt{V_t} dW_t + J_X dN_t \\ \log V_t &= \alpha + \beta F_t, \quad dF_t = \kappa F_t dt + dB_t + J_F dN_t - \nu\lambda dt. \end{aligned} \tag{4.3}$$

with $\mathbb{E}(dW_t dB_t) = \rho dt$, and $J_X \sim N(\mu_X, \sigma_X^2)$. We fix $\rho = -0.75$, $\mu_X = -0.02$, $\sigma_X = 0.05$, $\nu = 0.02$, $\sigma = 0.02$, $\lambda = 25$, and $\mu = 0.03$. This data generating process differs from (4.1) in

Table 1: Simple Linear Regression Results I

		Median Bias					
			Uncorrected			Corrected	
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	0.013	0.010	0.007	0.000	0.000	0.000
	1 min	0.021	0.017	0.012	0.000	0.000	0.000
b	15 sec	-0.053	-0.042	-0.033	0.000	0.000	-0.001
	1 min	-0.088	-0.070	-0.056	-0.003	-0.002	-0.002
		Median Absolute Error					
			Uncorrected			Corrected	
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	0.008	0.006	0.005	0.001	0.001	0.001
	1 min	0.013	0.011	0.008	0.002	0.002	0.002
b	15 sec	0.018	0.015	0.012	0.008	0.007	0.007
	1 min	0.027	0.023	0.019	0.016	0.016	0.015

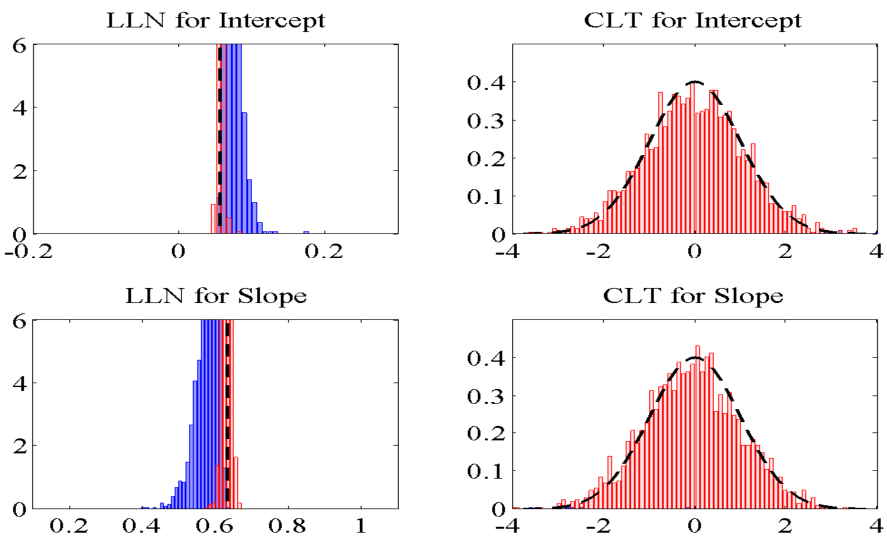
Note: In this table, we report the median bias and median absolute error of the biased and bias-corrected estimators in the simple linear regression of Section 4.1. T is chosen as 63 days, the sampling frequency ranges between $\Delta = 15$s and 60s, and $k_n = [\bar{\kappa}^{2/5}(\Delta)^{-2/5}]$.

Table 2: Simple Linear Regression Results II

		Nominal Level 10%					
		Uncorrected			Corrected		
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	100.00	100.00	99.95	14.50	13.20	13.95
	1 min	100.00	99.85	99.50	16.70	15.50	14.55
b	15 sec	100.00	99.95	99.65	12.40	12.10	12.15
	1 min	100.00	99.90	99.15	14.15	14.05	13.40
		Nominal Level 5%					
		Uncorrected			Corrected		
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	100.00	100.00	99.90	8.75	7.75	7.55
	1 min	99.95	99.80	99.40	10.05	8.90	7.50
b	15 sec	100.00	99.95	99.50	6.55	6.50	7.00
	1 min	100.00	99.90	98.95	7.85	8.10	6.95

Note: We report the rejection probabilities of 10% and 5% nominal level tests of the raw and two-step estimators in the simple linear regression of Section 4.1. T is chosen as 63 days, the sampling frequency ranges between $\Delta = 15$s and 60s, and $k_n = [\bar{\kappa}^{2/5}(\Delta)^{-2/5}]$.

Figure 1: Histograms for the Simple Linear Regression



Note: This figure compares the finite-sample distributions of estimators for slope and intercept in the simple linear regression model. The histograms in red are constructed using feasible the bias-corrected estimator, whereas those in blue are plotted using the naive estimator. The black dashed lines highlight the true values of parameters on the left panels, and the standard normal distribution on the right panels. We select 2000 Monte Carlo samples with $\Delta = 15$ s and $T = 63$ days. We choose $k_n = 172$ or $\bar{\kappa} = 1$ for estimation.

an important way in that the spot variance process is generated according to an exponential-OU model. We calculate the VIX based on the following model:

$$Y_t = a + bV_t^c + \varepsilon_t, \tag{4.4}$$

where $a = 0.0725$, $b = 3.8184$, $c = 0.9592$ and $\varepsilon_t \sim IID\mathcal{N}(0, 0.03^2)$ are calibrated to the market data. We refer to (4.4) as the "ABC" model for simplicity. This model is a parsimonious approximation to a fairly complicated nonlinear VIX pricing model implied by (4.3); see Section 5 for details. The parameter of interest is $\theta = (a, b, c)$.

We estimate the model (4.4) via nonlinear least square and conduct bias-correction using the two-step estimator. The results are organized in a similar manner as those in Section 4.1; see Figure 4.2, Tables 3 and 4. The results clearly show that the asymptotic theory works well when $\Delta = 15s$, but for $\Delta = 1$ min, the t-statistics associated with the two-step estimator tend to over-reject the null hypothesis for parameters b and c. This finding suggests that there is likely an weak identification issue for the joint estimation of b and c. This being said, the two-step estimator again shows evident improvement relative to its naive counterpart.

4.3 Specification test of affine VIX models

In view of the simulation results in subsection 4.2, in particular the weak identification issue in the case with $\Delta = 1$ min, we further examine the performance of the Anderson-Rubin confidence sets motivated by our empirical application.

We are interested in testing the null-hypothesis $H_0 : c = 1$ in the ABC model, i.e. the VIX is affine in the spot variance. Such tests can be conducted by examining whether the confidence sets $CS^{SW}_{n,1-\alpha}$, $CS^{SWSub}_{n,1-\alpha}$ with $\theta^{(1)} = c$ and $\theta^{(2)} = (a, b)$, $CS^{AS}_{n,1-\alpha}$, includes $c = 1$, when projected to the c-dimension. For comparison, we also consider t-tests based on the raw estimator with the "naive" standard error, as well as t-tests based on the two-step and the one-step (bias-corrected) estimators. In summary, we consider six tests based on: a) the raw estimator (raw); b) the two-step estimator (two-step); c) the one-step estimator (one-step); d) $CS^{SW}_{n,1-\alpha}$ (SW); e) $CS^{SW}_{n,1-\alpha}$ (SW-Sub); and f) $CS^{AS}_{n,1-\alpha}$ (AS).

We conduct simulation under the null hypothesis based on the affine jump-diffusion model as in Section 4.1. Table 5 shows finite-sample rejection probabilities of the aforementioned tests at nominal levels 1%, 5% and 10%. We summarize the results as follows. First, the raw test almost always falsely rejects the null hypothesis. Second, the two-step and one-step tests reject much less than the raw test, but still suffer from considerable over-rejection. The over-rejection is mitigated at higher sampling frequency. Third, tests robust to weak-identification, especially the AS test, perform quite well. We hence adopt the AS test in empirical applications.

Table 3: Nonlinear Regression Results I

Median Bias

		Uncorrected			Corrected		
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	-0.01	-0.01	-0.01	0.00	0.00	0.00
	1 min	-0.02	-0.02	-0.01	0.00	0.00	0.00
b	15 sec	-0.41	-0.34	-0.27	0.02	0.00	-0.02
	1 min	-0.64	-0.54	-0.43	0.05	0.02	-0.02
c	15 sec	-0.08	-0.07	-0.05	0.00	0.00	0.00
	1 min	-0.13	-0.11	-0.09	0.01	0.00	-0.01

Median Absolute Error

		Uncorrected			Corrected		
	Δ	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$	$\bar{\kappa} = 0.5$	$\bar{\kappa} = 1$	$\bar{\kappa} = 2$
a	15 sec	0.06	0.06	0.04	0.00	0.00	0.00
	1 min	0.12	0.10	0.08	0.01	0.01	0.01
b	15 sec	0.34	0.29	0.26	0.21	0.20	0.20
	1 min	0.52	0.47	0.42	0.51	0.48	0.46
c	15 sec	0.03	0.02	0.02	0.03	0.03	0.03
	1 min	0.04	0.04	0.04	0.06	0.06	0.06

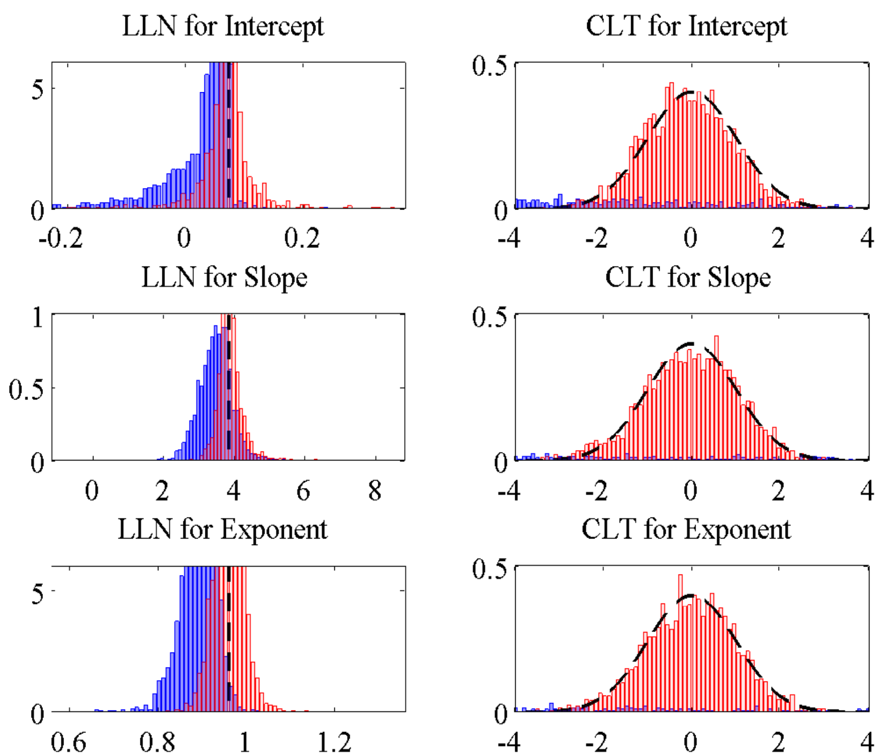
Note: In this table, we report the rejection probabilities of 10% and 5% level tests of the biased and bias-corrected estimators in the simple linear regression of Section 4.2. T is chosen as 63 days, the sampling frequency ranges between $\Delta = 15$s and 60s, and $k_n = [\bar{\kappa}^{2/5}(\Delta)^{-2/5}]$.

Table 4: Nonlinear Regression Results II

				Nominal Level 10			
			Uncorrected			Corrected	
	Δ	$\bar{\kappa}=0.5$	$\bar{\kappa}=1$	$\bar{\kappa}=2$	$\bar{\kappa}=0.5$	$\bar{\kappa}=1$	$\bar{\kappa}=2$
a	15 sec	96.80	94.25	93.05	10.70	9.35	11.20
	1 min	93.60	92.75	90.30	13.70	14.50	15.65
b	15 sec	98.35	97.45	96.55	14.35	12.05	12.45
	1 min	96.60	95.65	93.85	20.15	18.20	18.05
c	15 sec	98.60	97.55	96.45	10.85	10.30	11.20
	1 min	97.35	95.95	94.85	15.70	15.60	17.05
				Nominal Level 5			
			Uncorrected			Corrected	
	Δ	$\bar{\kappa}=0.5$	$\bar{\kappa}=1$	$\bar{\kappa}=2$	$\bar{\kappa}=0.5$	$\bar{\kappa}=1$	$\bar{\kappa}=2$
a	15 sec	95.90	93.20	91.45	5.35	5.05	5.55
	1 min	92.60	91.65	88.40	7.70	7.15	8.35
b	15 sec	98.25	97.00	95.75	7.45	6.95	7.45
	1 min	96.05	95.00	92.80	11.90	11.15	11.35
c	15 sec	98.15	97.20	96.15	6.00	5.95	6.25
	1 min	97.00	95.20	93.90	9.10	9.00	10.00

Note: In this table, we report the rejection probabilities of 10% and 5% level tests of the biased and bias-corrected estimators in the simple linear regression of Section 4.2. T is chosen as 63 days, the sampling frequency ranges between $\Delta = 15$s and 60s, and $k_n = [\bar{\kappa}^{2/5}(\Delta)^{-2/5}]$.

Figure 2: Histograms for the Nonlinear Regression



Note: This figure compares the finite-sample distributions of estimators for slope and intercept in the simple linear regression model. The histograms in red are constructed using feasible the bias-corrected estimator, whereas those in blue are plotted using the naive estimator. The black dashed lines highlight the true values of parameters on the left panels, and the standard normal distribution on the right panels. We select 2000 Monte Carlo samples with $\Delta = 15$ s and $T = 63$ days. We choose $k_n = 172$ or $\bar{\kappa} = 1$ for estimation.

Table 5: Comparison of Weak Identification Tests

	Raw	Two-Step	One-Step	SW	SW-Sub	AS
			$\Delta = 60$s			
1%	98.15	7.75	15.40	4.65	3.05	1.00
5%	98.7	17.90	29.00	13.2	10.7	5.00
10%	99.05	26.45	37.40	21.25	17.8	10.00
			$\Delta = 15$s			
1%	99.20	6.45	10.20	2.00	2.25	0.60
5%	99.30	14.80	20.80	8.15	7.45	4.20
10%	99.40	22.20	29.60	14.90	13.00	9.00

Note: In this table, we report the rejection probabilities (in percentage terms) of 1%, 5% and 10% nominal level tests introduced in Section 3.3. T is fixed at 63 days, the sampling frequency is chosen between $\Delta = 15$s and 60s, and $k_n = [0.5^{2/5}(\Delta)^{-2/5}]$.

5 Testing Option Pricing Models

5.1 Model Specification

We construct tests for option pricing models based on the relationship between the VIX and the unobserved spot variance. The dynamics of the index return under the risk-neutral measure, i.e. the "$\mathbb{Q}$ measure", is given as follows:

$$X_t = X_0 + (r-d)t + \int_0^t \sqrt{V_s} dW_s^Q + \int_0^t \int_{\mathbb{R}} j(x, V_s)\Big(N(ds, dx) - \nu(V_s, dx)ds\Big)$$

where N is a random counting measure with compensator ν that may depend on the spot variance. We assume there exists constants γ and η, such that

$$\int_{\mathbb{R}} j(x, v)^2 \nu(v, dx) = \gamma + \eta v.$$

This assumption is satisfied by many models in the empirical option pricing literature.

We consider two classes of risk-neutral models for the spot variance process which has been widely studied in empirical option pricing and financial econometrics. The first class, henceforth Type-I models, has the following risk-neutral dynamics:

$$V_t = V_0 + \int_0^t \kappa(\xi - V_s)ds + M_t^{\mathbb{Q}}$$

where $M^{\mathbb{Q}}$ is a martingale under the $\mathbb{Q}$-measure. Type-I models include those studied by Bakshi et al. (1997), Bates (2000), Pan (2002), Eraker (2004), Eraker et al. (2003), Broadie et al. (2007), Song and Xiu (2012), and Bates (2012), among others, where jumps may be driven by compound Poisson process with time-varying intensity or the CGMY process (Carr et al. (2003)). Type-I models also include non-Gaussian OU processes considered by Barndorff-Nielsen and Shephard (2001); see also Shephard (2005) for a collection of similar models. For models of this type, we can derive the following pricing formula for VIX.

Proposition 3 *Writing $Y_t = (VIX_t/100)^2$, we have for Type I models, that*

$$Y_t = a + bV_t, \tag{5.1}$$

where a and b only depend on model parameters, which are given in the Appendix 7.7,

Alternatively, the variance process of Type II has an exponential-affine structure shown below:

$$\log V_t = \alpha + \beta F_t, \quad F_t = F_0 + \int_0^t \kappa F_s ds + \sigma B_t^{\mathbb{Q}} + \int_0^t \int_{\mathbb{R}} f \cdot \Big(M(ds, df) - \mu(df)ds\Big).$$

which have been used in Huang and Tauchen (2005), Gonçalves and Meddahi (2009), and Barndorff-Nielsen et al. (2008).

Consequently, we have another pricing formula and also Tauchen and Todorov (2011).

Proposition 4 *For models of Type-II, $Y_t = (VIX_t/100)^2$ can be represented as*

$$Y_t = \frac{1}{\tau}\int_0^\tau \gamma + (\eta + 1)\exp\Big(\alpha + e^{\kappa u}(\log V_t - \alpha) + C(u)\Big)du, \tag{5.2}$$

where

$$C(u) = \int_0^u \psi(\beta e^{\kappa v})dv = \frac{1}{\kappa}\int_1^{e^{\kappa u}} \frac{\psi(\beta v)}{v}dv,$$

and $\psi(\cdot)$ is the cumulant function of the underlying Lévy process, that is,

$$\psi(u) = \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (e^{uf} - 1 - uf)\mu(df).$$

The ABC model can be considered as a parsimonious approximation of the VIX models of Type II. To see this, we let $\tau \to 0$, and consider the first order Taylor approximation of the Type

II pricing formula

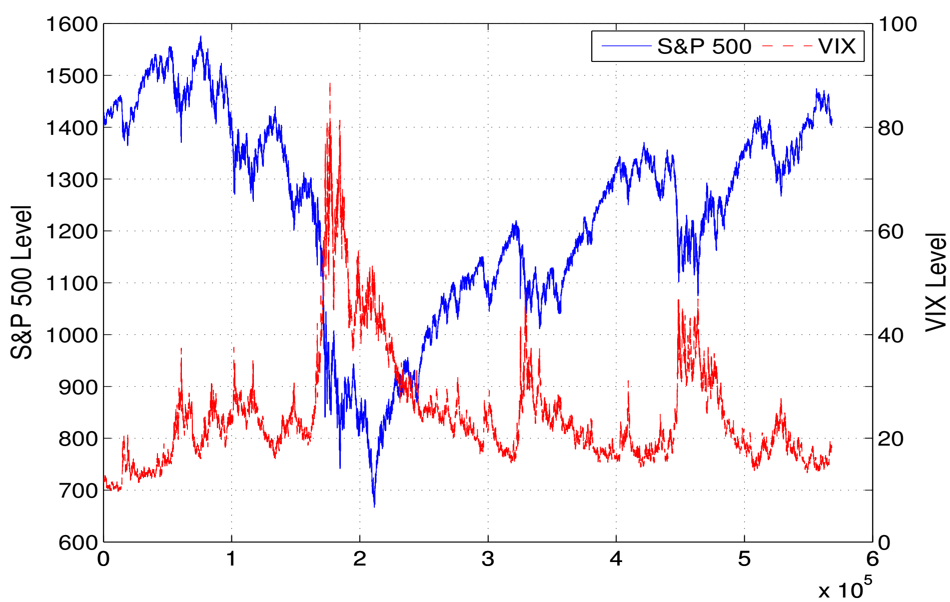
$$y_t \approx \gamma + (\eta + 1) \exp\left(\alpha + e^{\kappa\tilde{\tau}}(\log V_t - \alpha) + C(\tilde{\tau})\right) := a + b \cdot V^c \tag{5.3}$$

Numerically, we find these simple formulae good approximations of theoretical values of the VIX.

5.2 Data

We collect the intraday time series of the S&P 500 index and VIX ranging from Jan 2007 to September 2012, from the TickData Inc.. Figure 5.2 illustrates the data. The data are announced roughly every 15 seconds by the CBOE. We eliminate those days with trading hours shorter than regular trading hours, such as July 3rd, the day after Thanksgiving, etc, and end up with a sample of 1457 days. For each day, we synchronize the index and the VIX and obtain data pairs sampled at 1 min frequency. Finally, we divide and group the daily time series by quarter with a total of 23 quarters.

Figure 3: Time Series of the S&P 500 index and the VIX



Note: This figure reports the time series of data from Jan 1st, 2007 to Sep 30, 2012.

5.3 Empirical Implication

First, we estimate the affine VIX model (4.2) quarter-by-quarter, and report the estimates in Figure 5.3. The diurnal pattern of intraday spot variance are adjusted in estimation, following the standard procedure in the literature. The raw estimates are marked in red, whereas the black line plots the two-step estimates. Uniform confidence bands are shown in black dashed lines. Although most raw estimates fall into the uniform confidence bands, this should not be interpreted as insignificant difference, in that the theory and simulations suggest the considerable size distortion of the "naive" confidence bands.

Figure 5.3 exhibits statistically significant time variation in the parameter path across the entire sampling period, hence any parametric model of Type I with constant parameters cannot capture such time-varying linear relationship between the VIX and spot variance. Interestingly, the estimates of the intercept during the financial crisis (2008 Q3), the European (2010 Q2) and US Debt Crisis (2011 Q3), are significantly larger than average. This may indicate that additional risk factors beyond the stochastic volatility are missing in Type I models.

We further conduct tests for the specification of the affine VIX model. To do so, we nest the affine VIX model in the ABC model and test the null hypothesis $H_0 : c = 1$. Based on the simulation results, we use the AS test among the 6 alternatives. The testing strategy can be considered as augmenting the linear regression with an additional instrument, i.e., the derivative of the pricing function w.r.t the parameter c. The null hypothesis is rejected if the confidence sets for (a, b), with $c = 1$ fixed, is empty. At 5% nominal level, the null hypothesis is rejected in the aforementioned 3 "crisis" quarters. This finding further indicates that the affine VIX model is inadequate in capturing the relationship between the VIX and the spot variance. At 10% nominal level, the affine VIX model is also rejected in 2008 Q2.

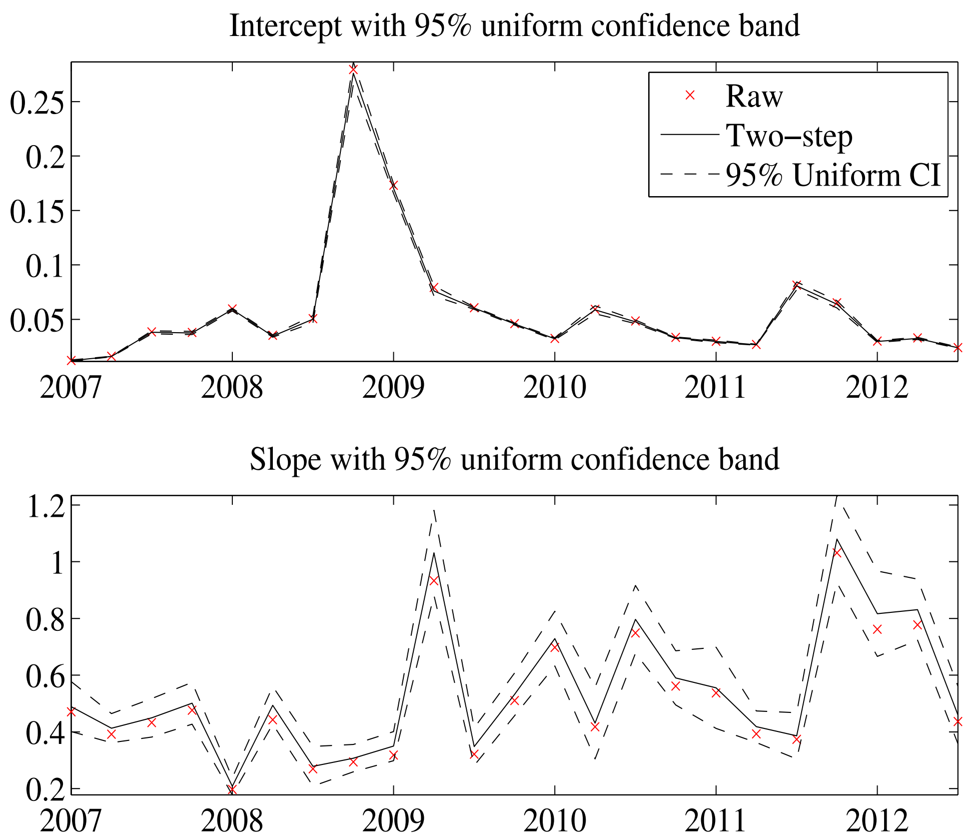
It may be worth mentioning that rejection may be due to the substantial variation in the volatility level over the crisis periods, which allows sufficient power for the AS test. Therefore, it is by no means safe to claim that the affine VIX model fits the remaining days. At least, such affine models suffer from the parameter stability issue.

6 Conclusion

References

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012), "Testing for jumps in noisy high frequency data," *Journal of Econometrics*, 168, 207–222.

Figure 4: Parameter Path in the Affine VIX Model



Note: This figure reports the quarterly estimates of (a, b) in the affine specification of the VIX. Raw estimates are marked in red, whereas the black line denotes the two-step estimates. 95% Uniform CI bands are plotted in black dashed lines. The sampling frequency is fixed at $\Delta = 1$ min, and $k_n = [\Delta^{-2/5}]$.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andrews, D. W. K. and Soares, G. (2010), "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical Performance of Alternative Option Pricing Models," *The Journal of Finance*, 52, 2003–2049.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), "Designing realized kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167–241.

Bates, D. (2012), "U.S. stock market crash risk, 1926-2010." *Journal of Financial Economics*, 105, 229–259.

Bates, D. S. (2000), "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

Broadie, M., Chernov, M., and Johannes, M. (2007), "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 62, 1453–1490.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003), "Stochastic volatility for Lévy processes," *Mathematical Finance*, 13, 345–342.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *The Journal of Finance*, 59, 1367–1404.

Eraker, B., Johannes, M. S., and Polson, N. (2003), "The Impact of Jumps in Equity Index Volatility and Returns," *The Journal of Finance*, 58, 1269–1300.

Gonçalves, S. and Meddahi, N. (2009), "Bootstrapping realized volatility," *Econometrica*, 77, 283–306.

Hall, A. (2005), *Generalized method of moments*, Oxford University Press.

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

Huang, X. and Tauchen, G. T. (2005), "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 4, 456–499.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., Podolskij, M., and Vetter, M. (2010), "Limit Theorems for Moving Averages of Discretized Processes Plus Noise," *Annals of Statistics, forthcoming.*

Jacod, J. and Protter, P. (2012), *Discretization of Processes*, Springer.

Jacod, J. and Rosenbaum, M. (2012), "Quarticity and Other Functionals of Volatility: Efficient Estimation," Tech. rep., Université de Paris-6.

Mykland, P. A. and Zhang, L. (2009), "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

Newey, W. K. and McFadden, D. L. (1994), *Large sample estimation and hypothesis testing*, Elsevier.

Novikov, A. (2003), "Martingales and first-passage times for Ornstein-Uhlenbeck process with a jump component," *Theory, Probability and Applications*, 48, 340–358.

Pan, J. (2002), "The jump-risk premia implicit in options: Evidence from an integrated time-series study," *Journal of Financial Economics*, 63, 3–50.

Shephard, N. (2005), *Stochastic volatility*, Oxford University Press.

Song, Z. and Xiu, D. (2012), "A Tale of Two Option Markets: State-Price Densities Implied from S&P 500 and VIX Option Prices," Tech. rep., Federal Reserve Board and University of Chicago.

Stock, J. H. and Wright, J. H. (2000), "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Tauchen, G. E. and Todorov, V. (2011), "Volatility Jumps," *Journal of Business & Economic Statistics*, 29, 356–371.

Todorov, V. and Bollerslev, T. (2010), “Jumps and betas: A new framework for disentangling and estimating systematic risks,” *Journal of Econometrics*, 157, 220 – 235.

Todorov, V. and Tauchen, G. (2011), “The Realized Laplace Transform of Volatility,” *Econometrica, Forthcoming.*

7 Proofs.

7.1 Notations and preliminary results

Throughout the proofs, we use K to denote a generic constant which may vary from line to line. A sequence of random variables X_n is $o_p(1)$ if $X_n \xrightarrow{\mathbb{P}} 0$; a sequence of random functions $X_n(\theta)$, $\theta \in \Theta$ is $o_{pu}(1)$ if $X_n(\theta) \xrightarrow{\mathbb{P}} 0$ uniformly in $\theta \in \Theta$. As is typical in this kind of problems, by a standard localization procedure, Assumption H1 can be strengthened into the following stronger version.

Assumption SH1. We have Assumption H1. The processes b_t and σ_t are bounded and $|\delta(\omega, t, z)| \leq \Gamma(z)$ for some deterministic function Γ on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} \Gamma(z)^r \lambda(dz) < \infty$. Moreover, $\tilde{b}_t$, $\tilde{\sigma}_t$ and $\tilde{\sigma}_t'$ are bounded and $|\tilde{\delta}(\omega, t, z)| \leq \tilde{\Gamma}(z)$ for some determistic function $\tilde{\Gamma}$ satisfying $\int_{\mathbb{R}} \tilde{\Gamma}(z)^2 \lambda(dz) < \infty$. The coefficients of Y^* and Z_t satisfy similar conditions for those of σ_t.

We collect several technical lemmas which are used repeatedly in the sequel.

Lemma 2 *Let $u \geq 1$ be a constant. Suppose (i) $(\tilde{Z}_t)_{t\geq 0}$ is càdlàg and takes value in some compact Euclidean set $\mathcal{K}$; (ii) $h : \mathcal{K} \times \mathcal{V} \times \Theta \mapsto \mathbb{R}$ is continuous and $\sup_{\tilde{z}\in\mathcal{K}} |h(\tilde{z}, v; \theta)| \leq C(1 + |v|^q)$ for some $q \geq 1$, $C > 0$ and all $v \in \mathcal{V}$; (iii) Assumption SH1 holds; (iv) $\varpi \geq (q-u)/(2q-r)$. Then*

(a) if $u = 1$, $\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta) - \int_0^T h(\tilde{Z}_s, V_s; \theta) ds = o_p(1)$ for each $\theta \in \Theta$;

(b) if we further assume $\sup_{\tilde{z}\in\mathcal{K}, \theta\in\Theta} \|\partial_\theta h(\tilde{z}, v; \theta)\| \leq C(1 + |v|^q)$, then (a) holds uniformly in $\theta \in \Theta$;

(c) if $u > 1$, $\Delta_n^u \sum_{i=0}^{[T/\Delta_n]-k_n} h(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta) = o_p(1)$ for each $\theta \in \Theta$.

Proof. (a) In this part, we fix $\theta \in \Theta$ and suppress the dependence of h on θ for notational simplicity. Without loss of generality, we can assume that h is positive (otherwise, consider the positive and the negative parts of h separately). We first prove the assertion under the assumption that h is bounded. We set, for $i \geq 1$,

$$\widehat{Z}_t = \tilde{Z}_{i\Delta_n} \text{ and } \widehat{V}_t = \widehat{V}_{i\Delta_n} \text{ for } t \in [(i-1)\Delta_n, i\Delta_n).$$

Observe that

$$\begin{aligned} & \mathbb{E}\left|\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - \int_0^T h(Z_s, V_s) ds\right| \\ \leq \quad & K k_n \Delta_n + \int_0^{([T/\Delta_n]-k_n)\Delta_n} \mathbb{E}\left|h\left(\widehat{Z}_s, \widehat{V}_s\right) - h(Z_s, V_s)\right| ds. \end{aligned}$$

Under Assumption SH1, we have $\widehat{V}_t \xrightarrow{\mathbb{P}} V_t$ for each $t \geq 0$ (Theorem 9.3.2, Jacod and Protter (2012)). Hence by the right continuity of $\tilde{Z}$, $(\widehat{Z}_s, \widehat{V}_s) \xrightarrow{\mathbb{P}} (\tilde{Z}_s, V_s)$; moreover, by the continuity of h, $h(\widehat{Z}_s, \widehat{V}_s) - h(\tilde{Z}_s, V_s) = o_p(1)$. Hence, by bounded convergence,

$$\int_0^{([T/\Delta_n]-k_n)\Delta_n} \mathbb{E}\left|h(\widehat{Z}_s, \widehat{V}_s) - h\left(\tilde{Z}_s, V_s\right)\right| ds \to 0,$$

yielding

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - \int_0^T h\left(\tilde{Z}_s, V_s\right) ds \xrightarrow{\mathbb{P}} 0 \text{ for bounded } h. \tag{7.1}$$

We now prove the assertion in part (a) with the boundedness condition on h relaxed. Let $\bar{\psi}$ be a C^∞ function $\mathbb{R}_+ \mapsto [0,1]$, with $1_{[1,\infty)}(x) \leq \bar{\psi}(x) \leq 1_{[1/2,\infty)}(x)$, and for $m \geq 1$, we set $\bar{\psi}_m(v) = \bar{\psi}(|v|/m)$, $\bar{\psi}'_m(v) = 1 - \bar{\psi}_m(v)$. We then define $h_m(\tilde{z}, v) = h(\tilde{z}, v)\bar{\psi}_m(v)$ and $h'_m(\tilde{z}, v) = h(\tilde{z}, v)\bar{\psi}'_m(v)$. Note that h'_m is bounded (for $\tilde{z} \in \mathcal{K}$, $v \in \mathcal{V}$) and continuous. By (7.1),

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h'_m\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - \int_0^T h'_m\left(\tilde{Z}_s, V_s\right) ds \xrightarrow{\mathbb{P}} 0.$$

Since V is bounded, $\int_0^T h'_m\left(\tilde{Z}_s, V_s\right) ds = \int_0^T h\left(\tilde{Z}_s, V_s\right) ds$ for m large enough. Hence, it remains to show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}\left|\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h_m\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right| = 0. \tag{7.2}$$

Note that for $m \geq 1$,

$$\begin{aligned} \left|h_m\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right| &= \left|h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right| \psi_m\left(\widehat{V}_{i\Delta_n}\right) \\ &\leq \sup_{\tilde{z}\in\mathcal{K}} \left|h\left(\tilde{z}, \widehat{V}_{i\Delta_n}\right)\right| \psi_m\left(\widehat{V}_{i\Delta_n}\right) \\ &\leq K\left(1 + \left|\widehat{V}_{i\Delta_n}\right|^q\right) \psi_m\left(\widehat{V}_{i\Delta_n}\right) \\ &\leq K\left|\widehat{V}_{i\Delta_n}\right|^q 1_{\left\{\widehat{V}_{i\Delta_n} > m/2\right\}}. \end{aligned} \tag{7.3}$$

Jacod and Protter (2012) show that ((9.4.7)), for some sequence $a_n \to 0$,

$$\mathbb{E}\left[|\widehat{V}_{i\Delta_n}|^q 1_{\left\{\widehat{V}_{i\Delta_n} > m/2\right\}}\right] \leq K m^{-q} + K \Delta_n^{1-q+\varpi(2q-r)} a_n. \tag{7.4}$$

Combining (7.3) and (7.4) and use $\varpi \geq (q-1) / (2q-r)$, we have

$$\mathbb{E}\left|\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h_m\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right| \leq Km^{-q} + O(a_n),$$

which further implies (7.2) and hence the assertion of part (a).

(b) By part (a), it remains to verify the stochastic equicontinuity of the random function $\rho_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)$. By the condition in part (b), $\sup_{\theta\in\Theta} \|\partial_\theta \rho_n(\theta)\| \leq K\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} (1+|\widehat{V}_{i\Delta_n}|^q) = O_p(1)$, where the equality is by part (a). The stochastic equicontinuity of $\rho_n(\theta)$ then follows a mean-value expansion and Theorem 21.9 in Davidson (1994).

(c) With the same notation in part (a) but with m fixed, it is easy to see that $\Delta_n^u \sum_{i=0}^{[T/\Delta_n]-k_n} h'_m(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}) = o_p(1)$. Moreover, we have

$$\mathbb{E}\left|\Delta_n^u \sum_{i=0}^{[T/\Delta_n]-k_n} h_m\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right| \leq K\Delta_n^{u-1} + K\Delta_n^{u-q+\varpi(2q-r)} a_n \to 0,$$

where the convergence is due to condition (iv). The assertion then readily follows. □

Below, we set

$$X'_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad \widehat{V}'_{i\Delta_n} = (k_n\Delta_n)^{-1} \sum_{j=0}^{k_n-1} \left(\Delta_{i+j}^n X'\right)^2$$

Lemma 3 *Suppose (i) $(\tilde{Z}_t)_{t\geq 0}$ is càdlàg and takes value in some compact Euclidean set $\mathcal{K}$; (ii) $h : \mathcal{K} \times \mathcal{V} \mapsto \mathbb{R}$ is continuous; (iii) h is continuously differetiable in v and $\sup_{\tilde{z}\in\mathcal{K}} |\partial_v h(\tilde{z}, v)| \leq C\left(1+|v|^{q-1}\right)$ for some $q \geq 1$, $C > 0$ and all $v \in \mathcal{V}$; (iv) Assumption SH1 holds. Then for some deterministic sequence $a_n \to 0$,*

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}\right)\right) = O_p\left(a_n \Delta_n^{(2q-r)\varpi+1-q}\right).$$

If we further have (v) $\varpi \geq (2q-1)/2(2q-r)$, then

$$\Delta_n^{1/2} \sum_{i=0}^{[T/\Delta_n]-k_n} \left(h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - h\left(\tilde{Z}_{i\Delta_n}, \widehat{V}'_{i\Delta_n}\right)\right) = o_p(1).$$

Proof of Lemma 3. By a mean-value expansion and condition (iii), for all $\tilde{z} \in \mathcal{K}$,

$$\begin{aligned}
&\mathbb{E}\left|h\left(\tilde{z}, \widehat{V}_{i\Delta_n}\right) - h\left(\tilde{z}, \widehat{V}'_{i\Delta_n}\right)\right| \\
&\quad \leq K\left(1 + \left|\widehat{V}'_{i\Delta_n}\right|^{q-1} + \left|\widehat{V}_{i\Delta_n} - \widehat{V}'_{i\Delta_n}\right|^{q-1}\right)\left|\widehat{V}_{i\Delta_n} - \widehat{V}'_{i\Delta_n}\right| \\
&\quad \leq K a_n \Delta_n^{(2q-r)\varpi+1-q},
\end{aligned}$$

for some sequence $a_n \to 0$; the second inequality follows Hölder's inequality, the standard estimate that $\mathbb{E}|\widehat{V}'_{i\Delta_n}|^{q'} \leq K_{q'}$ for any $q' \geq 0$ (note that X' is continuous), and Lemma 13.2.6 of Jacod and Protter (2012). The first assertion readily follows. The second assertion then follows condition (v) and $a_n \to 0$. □

The following result generalizes of the univariate case of Theorem 3.1 in Jacod and Rosenbaum (2012) by allowing for random test functions. The generalization is straightforward but tedious, and is left to the online supplement to this paper. For i, we set $\alpha_i^n = (\Delta_i^n X')^2 - V_{(i-1)\Delta_n}\Delta_n$.

Lemma 4 *Suppose (i) for any compact $\mathcal{K} \subseteq \mathcal{Z} \times \mathcal{V}$, the function $(z, v, \hat{v}) \mapsto h(z, v, \hat{v})$: $\mathcal{Z} \times \mathcal{V} \times \mathcal{V} \mapsto \mathbb{R}$ satisfies $\sup_{z \in \mathcal{K}} |\partial_{\hat{v}}^j h(z, v, \hat{v})| \leq C(1 + |v|^{q-j})$ for some constant $C > 0$, $q \geq 3$ and all $j = 0, 1, 2, 3$; (ii) for each $z \in \mathcal{Z}$, $v \in \mathcal{V}$, $h(z, v, v) = 0$; (iii) $\varpi \in [(2q-1)/2(2q-r), 1/2)$. We have the following*

$$\begin{aligned}
&\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n}\left(h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) - \frac{1}{k_n}\partial_{\hat{v}}^2 h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\widehat{V}_{i\Delta_n}^2\right) \\
&= \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_{\hat{v}} h\left(Z_{i\Delta_n}, V_{i\Delta_n}, V_{i\Delta_n}\right)\frac{1}{k_n}\sum_{j=1}^{k_n}\alpha_{i+j}^n + o_p\left(\Delta_n^{1/2}\right).
\end{aligned}$$

7.2 Proof of Theorem 1

Lemma 5 $\widehat{G}_n(\theta) - G(\theta) = o_{pu}(1)$ *and* $\widehat{Q}_n(\theta) - Q(\theta) = o_{pu}(1)$.

Proof. By localization, we can assume that (Y^*, Z) takes value in $\mathcal{K} = \mathcal{K}_{\mathcal{Y}} \times \mathcal{K}_{\mathcal{Z}}$ where $\mathcal{K}_{\mathcal{Y}}$ and $\mathcal{K}_{\mathcal{Z}}$ are compact, and A_t bounded. Denote

$$\widehat{G}_n^*(\theta) = \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} g(Y_{i\Delta_n}^*, Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta), \quad \theta \in \Theta.$$

Applying Lemma 2(b) with $\tilde{Z} = (Y^*, Z)$, we have $\widehat{G}_n^*(\theta) - G(\theta) = o_{pu}(1)$.

We now show that $\widehat{G}_n(\theta) - \widehat{G}_n^*(\theta) = o_{pu}(1)$. By arguing component by component, we can and shall assume that $J = k = 1$; hence $\widehat{G}_n(\theta) - \widehat{G}_n^*(\theta) = -\Delta_n \sum_{i=1}^{[T/\Delta_n]-k_n+1} d_1(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)\varepsilon_{1,i\Delta_n}$.

By Assumption H3,

$$\mathbb{E}\left[\left(\widehat{G}_n(\theta)-\widehat{G}_n^*(\theta)\right)^2\middle|\mathcal{F}\right] = \Delta_n^2\sum_{i=0}^{[T/\Delta_n]-k_n} d_1(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta)^2A_{i\Delta_n}^2 \xrightarrow{\mathbb{P}} 0,$$

where the convergence follows Lemma 2(c) with $q=2p$ and $u=2$. Hence, $\widehat{G}_n(\theta)-\widehat{G}_n^*(\theta)\xrightarrow{\mathbb{P}}0$ for each $\theta\in\Theta$. By a mean-value expansion, we have, for θ, $\theta'\in\Theta$,

$$\left|\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n} d_1(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta')\varepsilon_{1,i\Delta_n}-\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n} d_1(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta)\varepsilon_{1,i\Delta_n}\right| \leq \left(\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\sup_{z\in\mathcal{K}_{\mathcal{Z}},\theta\in\Theta}\left\|\partial_\theta d_1(z,\widehat{V}_{i\Delta_n};\theta)\right\||\varepsilon_{1,i\Delta_n}|\right)\|\theta'-\theta\|.$$

Observe that

$$\mathbb{E}\left[\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\sup_{z\in\mathcal{K}_{\mathcal{Z}},\theta\in\Theta}\left\|\partial_\theta d(z,\widehat{V}_{i\Delta_n};\theta)\right\||\varepsilon_{1,i\Delta_n}|\middle|\mathcal{F}\right] \leq K\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\sup_{z\in\mathcal{K}_{\mathcal{Z}},\theta\in\Theta}\left\|\partial_\theta d(z,\widehat{V}_{i\Delta_n};\theta)\right\|.$$

By Lemma 2(a), the term on the majorant side of the above display is $O_p(1)$, so is $\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\sup_{z\in\mathcal{K}_{\mathcal{Z}},\theta\in\Theta}\|\partial_\theta d(z,\widehat{V}_{i\Delta_n};\theta)\||\varepsilon_{i\Delta_n}|$. Therefore, $\widehat{G}_n(\theta)-\widehat{G}_n^*(\theta)$ is stochastically equicontinuous. It readily follows that $\widehat{G}_n(\theta)-\widehat{G}_n^*(\theta)=o_{pu}(1)$. The first assertion of the lemma is then obvious. Since $\Xi_n=O_p(1)$ and $\sup_{\theta\in\Theta}\|G(\theta)\|=O_p(1)$, the second assertion follows directly from the first assertion. □

Proof of Theorem 1. By localization, we can assume that Y^*, Z and V are bounded. Since g is continuous, $g(Y_s^*,Z_s,V_s;\theta)$ is uniformly bounded. By bounded convergence, $G(\cdot)$ is continuous, so is $Q(\cdot)$. By a standard argument (e.g. Theorem 2.1, Newey and McFadden (1994)), the assertion of the theorem readily follows Assumptions GMM1 and Lemma 5. □

7.3 Proofs in Section 2.3

Lemma 6 *For each $1 \le j \le J$, we have*

$$k_n^{-1}\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \gamma_j\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n};\theta_0\right) = k_n^{-1}\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \gamma_j\left(Z_{i\Delta_n}, V_{i\Delta_n};\theta_0\right) + o_p\left(\Delta_n^{1/2}\right).$$

Proof of Lemma 6. By localization, we can assume that Z and V are bounded. Observe that $\sup_{z\in\mathcal{K}} |\partial_v h_{m,j}(z,v)| \le K(|v| + |v|^{2p-1})$. By Lemma 3,

$$k_n^{-1}\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h_{m,j}(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}) = k_n^{-1}\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h_{m,j}(Z_{i\Delta_n}, \widehat{V}'_{i\Delta_n}) + o_p\left(k_n^{-1}\Delta_n^{1/2}\right). \tag{7.5}$$

Observe that for all $z \in \mathcal{K}$,

$$\left|h_{m,j}\left(z, \widehat{V}'_{i\Delta_n}\right) - h_{m,j}\left(z, V_{i\Delta_n}\right)\right| \le K\left(1 + \left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right|^{2p-1}\right)\left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right|.$$

Note that $\widehat{V}'_{i\Delta_n} - V_{i\Delta_n} = \bar{\alpha}_i^n + \bar{\beta}_i^n$ and for $q \ge 2$, $\mathbb{E}\left|\bar{\alpha}_i^n\right|^q \le K_q k_n^{-q/2}$, $\mathbb{E}\left|\bar{\beta}_i^n\right|^q \le K_q k_n \Delta_n$. Hence,

$$\mathbb{E}\left|h_{m,j}(Z_{i\Delta_n}, \widehat{V}'_{i\Delta_n}) - h_{m,j}\left(Z_{i\Delta_n}, V_{i\Delta_n}\right)\right| \le K\left(k_n^{-1/2} + \sqrt{k_n\Delta_n}\right).$$

Therefore,

$$\frac{\Delta_n}{k_n} \sum_{i=0}^{[T/\Delta_n]-k_n} \left|h_{m,j}(Z_{i\Delta_n}, \widehat{V}'_{i\Delta_n}) - h_{m,j}\left(Z_{i\Delta_n}, V_{i\Delta_n}\right)\right| = O_p\left(\frac{1}{k_n^{3/2}} + \frac{\Delta_n^{1/2}}{k_n^{1/2}}\right). \tag{7.6}$$

Combining (7.5) and (7.6), we readily derive the first assertion by using $k_n^3\Delta_n \to \infty$. The second assertion follows the Riemann approximation and $k_n\Delta_n^{1/2} \xrightarrow{\mathbb{P}} 0$. $\square$

Below, we set $\xi_n = (\xi_{1,n}, \ldots, \xi_{J,n})$ where for $1 \le j \le J$,

$$\begin{aligned} \xi_{j,n} &= \xi_{1,j,n} + \xi_{2,j,n}, \\ \xi_{1,j,n} &= \Delta_n^{-1/2} \sum_{i=0}^{[T/\Delta_n]-k_n} \phi_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) \frac{1}{k_n} \sum_{j=1}^{k_n} \alpha_{i+j}^n \end{aligned}$$

$$\xi_{2,j,n} = -\Delta_n^{1/2} \sum_{i=0}^{[T/\Delta_n]-k_n} d_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) \varepsilon_{j,i\Delta_n}.$$

Proposition 5 $\widehat{G}_n(\theta_0) = \Delta_n^{1/2}\xi_n + k_n^{-1}\widehat{B}_n(\theta_0) + o_p\left(\Delta_n^{1/2}\right).$

Proof. We fix some $j \in \{1, \ldots, J\}$. Let $h(z, v, \hat{v}) = d_j(z, \hat{v}; \theta_0)(f_j(z, \hat{v}; \theta_0) - f_j(z, v; \theta_0))$. We decompose

$$\begin{aligned}
\widehat{G}_{j,n}(\theta_0) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) \\
&\quad -\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} d_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) \varepsilon_{j,i\Delta_n} + R_{n,1} \\
R_{n,1} &\equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(d_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) - d_j\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right) \varepsilon_{j,i\Delta_n}.
\end{aligned}$$

By Lemma 4,

$$\begin{aligned}
&\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) \\
= &\frac{\Delta_n}{k_n} \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_{\hat{v}}^2 h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) \widehat{V}_{i\Delta_n}^2 \\
&+\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_{\hat{v}} h\left(Z_{i\Delta_n}, V_{i\Delta_n}, V_{i\Delta_n}\right) \frac{1}{k_n} \sum_{j=1}^{k_n} \alpha_{i+j}^n + o_p\left(\Delta_n^{1/2}\right).
\end{aligned}$$

Observe that

$$\begin{aligned}
\partial_{\hat{v}} h(z, v, \hat{v}) &= \partial_v d_j(z, \hat{v}; \theta_0)(f_j(z, \hat{v}; \theta_0) - f_j(z, v; \theta_0)) \\
&\quad + d_j(z, \hat{v}; \theta_0) \partial_v f_j(z, \hat{v}; \theta_0). \\
\partial_{\hat{v}}^2 h(z, v, \hat{v}) &= \partial_v^2 d_j(z, \hat{v}; \theta_0)(f_j(z, \hat{v}; \theta_0) - f_j(z, v; \theta_0)) \\
&\quad + 2\partial_v d_j(z, \hat{v}; \theta_0) \partial_v f_j(z, \hat{v}; \theta_0) + d_j(z, \hat{v}; \theta_0) \partial_v^2 f_j(z, \hat{v}; \theta_0).
\end{aligned}$$

Hence,

$$\frac{\Delta_n}{k_n} \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_{\hat{v}}^2 h\left(Z_{i\Delta_n}, V_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right) \widehat{V}_{i\Delta_n}^2 = k_n^{-1}\widehat{B}_n(\theta_0) + R_{n,2},$$

where

$$R_{n,2} \equiv \frac{\Delta_n}{k_n} \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_v^2 d_j \left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta_0 \right) \times \left(f_j(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta_0) - f_j\left(Z_{i\Delta_n}, V_{i\Delta_n}; \theta_0\right) \right) \widehat{V}_{i\Delta_n}^2.$$

Moreover,

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_{\hat{v}} h\left(Z_{i\Delta_n}, V_{i\Delta_n}, V_{i\Delta_n}\right) \frac{1}{k_n} \sum_{j=1}^{k_n} \alpha_{i+j}^n = \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \phi_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) \frac{1}{k_n} \sum_{j=1}^{k_n} \alpha_{i+j}^n$$

It remains to show that $R_{n,j} = o_p\left(\Delta_n^{1/2}\right)$ for $j = 1, 2$; this is the task below.

By Assumption H3,

$$\begin{aligned}
&\mathbb{E}\left[\left(\Delta_n^{-1/2} R_{n,1}\right)^2 \middle| \mathcal{F} \right] \\
&= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(d_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) - d_j\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right)^2 A_{jj,i\Delta_n} \\
&\leq K\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(d_j\left(Z_{i\Delta_n}, V_{i\Delta_n}\right) - d_j\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}\right)\right)^2 \\
&\xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

where the convergence is by Lemma 2(a). Hence, $R_{n,1} = o_p(\Delta_n^{1/2})$.

Let $R'_{n,2}$ be defined as $R_{n,2}$ but with $\widehat{V}'_{i\Delta_n}$ in place of $\widehat{V}_{i\Delta_n}$. By Lemma 3, $R_{n,2} - R'_{n,2} = o_p(\Delta_n^{1/2})$. Note that

$$\begin{aligned}
&\left| \partial_v^2 d_j \left(Z_{i\Delta_n}, \widehat{V}'_{i\Delta_n}; \theta_0 \right) \left(f_j(Z_{i\Delta_n}, \widehat{V}'_{i\Delta_n}; \theta_0) - f_j\left(Z_{i\Delta_n}, V_{i\Delta_n}; \theta_0\right)\right) \widehat{V}'^2_{i\Delta_n} \right| \\
\leq\ & K\left(1 + \left|\widehat{V}'_{i\Delta_n}\right|^{p-2}\right)\left(1 + \left|\widehat{V}'_{i\Delta_n}\right|^{p-1}\right)\left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right| \widehat{V}'^2_{i\Delta_n} \\
\leq\ & K\left(1 + \left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right|^{2p-1}\right)\left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right|.
\end{aligned}$$

Since for any $q \geq 0$, $\mathbb{E}\left|\widehat{V}'_{i\Delta_n} - V_{i\Delta_n}\right|^q \leq K_q k_n^{-q/2}$, we derive $R'_{n,2} = O_p(k_n^{-3/2}) = o_p(\Delta_n^{1/2})$. Hence, $R_{n,2} = o_p(\Delta_n^{1/2})$. This finishes the proof. □

Proposition 6 *The variables* $((\xi_{1,j,n})_{1\leq j\leq J}, (\xi_{2.j,n})_{1\leq j\leq J})$ *converges stably in law to variables*

$\left((\tilde{\xi}_{1,j})_{1\leq j\leq J},(\tilde{\xi}_{2,j})_{1\leq j\leq J}\right)$ *defined on an extension* $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$ *of the space* $(\Omega,\mathcal{F},\mathbb{P})$*, which conditionally on* $\mathcal{F}$*, are centered Gaussian with variance-covariance given as follows: for* $1\leq j,l\leq J$*,*

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\tilde{\xi}_{1,j}\tilde{\xi}_{1,l}^{\intercal}\middle|\,\mathcal{F}\right] &= 2\int_0^T \phi_j\left(Z_s,V_s;\theta_0\right)\phi_l\left(Z_s,V_s;\theta_0\right)^{\intercal}V_s^2ds\\
\widetilde{\mathbb{E}}\left[\tilde{\xi}_{2,j}\tilde{\xi}_{2,l}^{\intercal}\middle|\,\mathcal{F}\right] &= \int_0^T d_j\left(Z_s,V_s;\theta_0\right)d_l\left(Z_s,V_s;\theta_0\right)^{\intercal}A_{jl,s}ds\\
\widetilde{\mathbb{E}}\left[\tilde{\xi}_{1,j}\tilde{\xi}_{2,l}^{\intercal}\middle|\,\mathcal{F}\right] &= 0.
\end{aligned}$$

Proof. For notational simplicity, we set $\phi_{j,i}^n=\phi_j\left(Z_{i\Delta_n},V_{i\Delta_n};\theta_0\right)$ and $d_{j,i}^n=d_j\left(Z_{i\Delta_n},V_{i\Delta_n};\theta_0\right)$. We can rewrite

$$\begin{aligned}
\xi_{1,j,n} &= \sum_{i=0}^{[T/\Delta_n]\ -k_n}\phi_{j,i}^n\frac{1}{k_n\Delta_n^{1/2}}\sum_{k=1}^{k_n}\alpha_{i+k}^n\\
&= \sum_{i=k_n}^{[T/\Delta_n]\ -k_n}\left(\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{j,i-k}^n\right)\Delta_n^{-1/2}\alpha_i^n+O_p\left(k_n\Delta_n^{1/2}\right).
\end{aligned}$$

As shown in Jacod and Rosenbaum (2012) (see (4.12) and Lemma 4.2), $\mathbb{E}\left|\left[\left.\alpha_i^n\right|\mathcal{F}_{(i-1)\Delta_n}\right]\right|\leq K\Delta_n^{3/2}(\Delta_n^{1/2}+\eta_i^n)$ for some variables η_i^n satisfying $\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\eta_i^n\xrightarrow{\mathbb{P}}0$. It readily follows that

$$\sum_{i=k_n}^{[T/\Delta_n]\ -k_n}\left(\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{j,i-k}^n\right)\Delta_n^{-1/2}\left|\mathbb{E}\left[\left.\alpha_i^n\right|\mathcal{F}_{(i-1)\Delta_n}\right]\right|\xrightarrow{\mathbb{P}}0.$$

We also observe that

$$\begin{aligned}
&\sum_{i=k_n}^{[T/\Delta_n]\ -k_n}\left(\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{j,i-k}^n\right)\left(\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{l,i-k}^n\right)^{\intercal}\mathbb{E}\left[\Delta_n^{-1}\left(\alpha_i^n\right)^2\middle|\,\mathcal{F}_{(i-1)\Delta_n}\right]\\
&\xrightarrow{\mathbb{P}}2\int_0^T\phi_j\left(Z_s,V_s;\theta_0\right)\phi_l\left(Z_s,V_s;\theta_0\right)^{\intercal}V_s^2ds,
\end{aligned}$$

Moreover, $\xi_{2,j,n}=-\Delta_n^{1/2}\sum_{i=k_n}^{[T/\Delta_n]-k_n}d_{j,i-1}^n\varepsilon_{j,(i-1)\Delta_n}+o_p\left(1\right)$ and

$$\begin{aligned}
&\Delta_n\sum_{i=k_n}^{[T/\Delta_n]\ -k_n}(d_{j,i-1}^n)(d_{l,i-1}^n)^{\intercal}\mathbb{E}\left[\varepsilon_{j,(i-1)\Delta_n}\varepsilon_{l,(i-1)\Delta_n}|\mathcal{F}_{(i-1)\Delta_n}\right]\\
&\xrightarrow{\mathbb{P}}\int_0^T d_j\left(Z_s,V_s;\theta_0\right)d_l\left(Z_s,V_s;\theta_0\right)^{\intercal}A_{jl,s}ds.
\end{aligned}$$

By Assumption H3, $\mathbb{E}\left[\alpha_i^n\varepsilon_{j,(i-1)\Delta_n}\right]=0$.

It is easily seen that

$$\sum_{i=k_n}^{[T/\Delta_n]\ -k_n} \mathbb{E}\left[\left\|\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{j,i-k}^n\Delta_n^{-1/2}\alpha_i^n\right\|^4\middle|\mathcal{F}_{(i-1)\Delta_n}\right] \leq K\Delta_n\to 0,$$

$$\sum_{i=k_n}^{[T/\Delta_n]\ -k_n} \mathbb{E}\left[\left\|d_j\left(Z_s,V_s;\theta_0\right)\Delta_n^{1/2}\varepsilon_{j,(i-1)\Delta_n}\right\|^4\middle|\mathcal{F}_{(i-1)\Delta_n}\right] \leq K\Delta_n\to 0.$$

By routine manipulation, for $N = W$ or N being an arbitrary bounded martingale on $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$,

$$\sum_{i=k_n}^{[T/\Delta_n]\ -k_n}\left(\frac{1}{k_n}\sum_{k=1}^{k_n}\phi_{j,i-k}^n\right)\Delta_n^{-1/2}\mathbb{E}\left[\alpha_i^n\Delta_i^nN\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\xrightarrow{\mathbb{P}}0$$

$$\sum_{i=k_n}^{[T/\Delta_n]\ -k_n} d_{j,i-1}^n\mathbb{E}\left[\varepsilon_{j,(i-1)\Delta_n}\Delta_i^nN\middle|\mathcal{F}_{(i-1)\Delta_n}\right]=0.$$

The assertion then follows from Theorem IX.7.28 in Jacod and Protter (2012). □

Proof of Proposition 1. (a) The assertion follows a similar argument as Lemma 5(a). (b) By Lemma 2, $\widehat{B}_n\left(\theta_0\right)\xrightarrow{\mathbb{P}}B$. By Proposition 5 and Proposition 6, $\widehat{G}_n\left(\theta_0\right)=k_n^{-1}\widehat{B}_n\left(\theta_0\right)+O_p(\Delta_n^{1/2})$. Since $k_n\Delta_n^{1/2}\to 0$, we derive $k_n\widehat{G}_n\left(\theta_0\right)\xrightarrow{\mathbb{P}}B$ as desired. Part (c) follows directly from part (a,b), (2.4) and Theorem 1. □

Proof of Lemma 1. By a component-wise argument, we assume that $k=\dim\left(\theta\right)=1$ without loss. Observe that

$$\begin{aligned}
\partial_\theta\widehat{H}_n\left(\theta\right) &= \Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\partial_\theta^2 g\left(Y_{i\Delta_n},Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)\\
&= \Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\partial_\theta^2 d_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)\left(f_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)-Y_{1,i\Delta_n}\right)\\
&\quad +2\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n}\partial_\theta d_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)\partial_\theta f_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)\\
&\quad +\Delta_n\sum_{i=0}^{[T/\Delta_n]-k_n} d_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)\partial_\theta^2 f_1\left(Z_{i\Delta_n},\widehat{V}_{i\Delta_n};\theta\right)
\end{aligned}$$

Along with the same line of argument as in the proof of Lemma 5, we can show that $\sup_{\theta\in\Theta}\|\partial_\theta\widehat{H}_n\left(\theta\right)\| = O_p\left(1\right)$. By the mean value theorem, $|\widehat{H}_n\left(\theta_n^H\right)-\widehat{H}_n\left(\hat{\theta}_n\right)| \leq$

$\sup_{\theta\in\Theta} |\partial_\theta \widehat{H}_n(\theta)| \cdot |\theta_n^H - \hat{\theta}_n|$. By Proposition 1(c), $\theta_n^H - \hat{\theta}_n = O_p(k_n^{-1})$. The assertion in part (a) readily follows.

By part (a), $\widehat{H}_n(\hat{\theta}_n) = \widehat{H}_n(\theta_0) + O_p(k_n^{-1})$. By a similar argument, $\widehat{H}_n^*(\theta_n^H) - \widehat{H}_n^*(\theta_0) = O_p(k_n^{-1})$. To show the assertion in part (b), it remains to show that $\widehat{H}_n(\theta_0) - \widehat{H}_n^*(\theta_0) = O_p\left(k_n^{-1}\right)$. By definition,

$$\widehat{H}_n(\theta_0) - \widehat{H}_n^*(\theta_0) = -\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \partial_\theta d_1\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n};\theta_0\right)\varepsilon_{1,i\Delta_n}.$$

Hence, by Lemma 2,

$$\begin{aligned}\mathbb{E}\left[k_n^2\left(\widehat{H}_n(\theta_0) - \widehat{H}_n^*(\theta_0)\right)^2\middle|\mathcal{F}\right] &= k_n^2\Delta_n^2 \sum_{i=0}^{[T/\Delta_n]-k_n}\left(\partial_\theta d_1\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n};\theta_0\right)\right)^2 A_{i\Delta_n}\\ &\leq k_n^2\Delta_n \cdot O_p(1).\end{aligned}$$

By (**??**), we readily derive $\widehat{H}_n(\theta_0) - \widehat{H}_n^*(\theta_0) = O_p(k_n^{-1})$. This finishes the proof of part (b). □

Proof of Theorem 2. By Lemma 2,

$$\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \sup_{\theta\in\Theta_0} \|\partial_\theta\gamma(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n};\theta)\| = O_p(1). \tag{7.7}$$

By assumption, $\tilde{\theta}_n - \theta_0 = O_p\left(k_n^{-1}\right)$. Hence,

$$\widehat{B}_n(\tilde{\theta}_n) = \widehat{B}_n(\theta_0) + O_p\left(k_n^{-1}\right). \tag{7.8}$$

Recall (2.4). To simplify notations, we set $\widehat{M}_n \equiv -(\widehat{H}_n(\hat{\theta}_n)^\intercal \Xi_n \widehat{H}_n(\bar{\theta}_n))^{-1}\widehat{H}_n(\hat{\theta}_n)^\intercal\Xi_n$ and $\widetilde{M}_n = -(\widetilde{H}_n^\intercal\Xi_n\widetilde{H}_n)^{-1}\widetilde{H}_n^\intercal\Xi_n$. Since $\hat{\theta}_n - \theta_0 = O_p\left(k_n^{-1}\right)$, we also have $\hat{\theta}_n - \bar{\theta}_n = O_p\left(k_n^{-1}\right)$. By Lemma 1, $\widehat{H}_n(\hat{\theta}_n) - \widehat{H}_n(\bar{\theta}_n) = O_p\left(k_n^{-1}\right)$. Since $\widetilde{H}_n = \widehat{H}_n(\hat{\theta}_n) + O_p\left(k_n^{-1}\right)$, we have $\widehat{M}_n - \widetilde{M}_n = O_p\left(k_n^{-1}\right)$. Therefore,

$$\begin{aligned}\hat{\theta}_n^* - \theta_0 &= \hat{\theta}_n - \theta_0 - \frac{1}{k_n}\widetilde{M}_n\widehat{B}_n(\tilde{\theta}_n)\\ &= \widehat{M}_n\widehat{G}_n(\theta_0) - \frac{1}{k_n}\widetilde{M}_n\widehat{B}_n(\tilde{\theta}_n)\\ &= \left(\widehat{M}_n - \widetilde{M}_n\right)\widehat{G}_n(\theta_0) + \widetilde{M}_n\left(\widehat{G}_n(\theta_0) - \frac{1}{k_n}\widehat{B}_n(\tilde{\theta}_n)\right)\\ &= O_p\left(k_n^{-2}\right) + \widetilde{M}_n\left(\widehat{G}_n(\theta_0) - \frac{1}{k_n}\widehat{B}_n(\tilde{\theta}_n)\right).\end{aligned} \tag{7.9}$$

By Proposition 5 and (7.8),

$$\begin{aligned}
&\widehat{G}_n(\theta_0) - \frac{1}{k_n}\widehat{B}_n(\tilde{\theta}_n) \\
&\quad = \Delta_n^{1/2}\xi_n + \frac{1}{k_n}\left(\widehat{B}_n(\theta_0) - \widehat{B}_n(\tilde{\theta}_n)\right) + o_p\left(\Delta_n^{1/2}\right) \\
&\quad = \Delta_n^{1/2}\xi_n + O_p\left(k_n^{-2}\right) + o_p\left(\Delta_n^{1/2}\right).
\end{aligned} \tag{7.10}$$

By Proposition 1(a), $\widehat{M_n} \xrightarrow{\mathbb{P}} M$. Hence, $\widetilde{M_n} \xrightarrow{\mathbb{P}} M$. Observing $k_n^{-2} = o(\Delta_n^{1/2})$, by (7.9) and (7.10), we deduce

$$\Delta_n^{-1/2}\left(\theta_n^* - \theta_0\right) = M\xi_n + o_p\left(1\right). \tag{7.11}$$

The assertion then follows Proposition 6. □

7.4 Proofs in Section 2.4

Proof of Theorem 3. The key is to show that

$$\sup_{\theta\in\Theta}\left\|\widehat{S}_n\left(\theta\right) - S\left(\theta\right)\right\| \xrightarrow{\mathbb{P}} 0. \tag{7.12}$$

By Lemma 2, $\widehat{S}_{2,n}\left(\theta\right) \xrightarrow{\mathbb{P}} S_2(\theta)$ uniformly. It remains to show $\widehat{S}_{1,n}\left(\theta\right) \xrightarrow{\mathbb{P}} S_1(\theta)$ uniformly. Fix $j, l \in \{1, \ldots, J\}$. Denote $\hat{d}_{j,i}\left(\theta\right) = d_j(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)$ and $d_{j,s}\left(\theta\right) = d_j\left(Z_s, V_s; \theta\right)$ for notational simplicity. Let $U_n\left(\theta\right)$ and $U\left(\theta\right)$ be respectively the (j, l) blocks of $\widehat{S}_{1,n}\left(\theta\right)$ and $S_1(\theta)$, i.e.

$$\begin{aligned}
U_n\left(\theta\right) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}\left(\theta\right)\hat{d}_{l,i}\left(\theta\right)^\intercal \hat{\varepsilon}_{j,i\Delta_n}\left(\theta\right)\hat{\varepsilon}_{l,i\Delta_n}\left(\theta\right) \\
U\left(\theta\right) &= \int_0^T d_{j,s}\left(\theta\right) d_{l,s}\left(\theta\right)^\intercal A_{jl,s} ds.
\end{aligned}$$

It remains to show that $U_n\left(\theta\right) \xrightarrow{\mathbb{P}} U\left(\theta\right)$ uniformly. By arguing component by component, we can assume that $k_j = k_l = 1$ without loss of generality. We set

$$\begin{aligned}
U_n'\left(\theta\right) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}\left(\theta\right)\hat{d}_{l,i}\left(\theta\right)\varepsilon_{j,i\Delta_n}\varepsilon_{l,i\Delta_n} \\
U_n''\left(\theta\right) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}\left(\theta\right)\hat{d}_{l,i}\left(\theta\right) A_{jl,i\Delta_n}.
\end{aligned}$$

By Lemma 2, we derive $U_n''\left(\theta\right) \xrightarrow{\mathbb{P}} U\left(\theta\right)$ uniformly in $\theta \in \Theta$. By Assumption E,

$\mathbb{E}\left[U_n'(\theta) - U_n''(\theta)\right| \mathcal{F}] = 0$ and

$$\begin{aligned} & \mathbb{E}\left[\left(U_n'(\theta) - U_n''(\theta)\right)^2 \middle| \mathcal{F}\right] \\ = \;& \Delta_n^2 \sum_{i=0}^{[T/\Delta_n]-k_n} \left(\hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\right)^2 \mathbb{E}\left[\left(\varepsilon_{j,i\Delta_n}\varepsilon_{l,i\Delta_n} - A_{jl,i\Delta_n}\right)^2 \middle| \mathcal{F}\right] \\ \leq \;& K\Delta_n^2 \sum_{i=0}^{[T/\Delta_n]-k_n} \left(\hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\right)^2 \\ = \;& o_p(1). \end{aligned}$$

It readily follows that $U_n'(\theta) - U_n''(\theta) \xrightarrow{\mathbb{P}} 0$ for each $\theta \in \Theta$. By a similar argument as in Lemma 5(a), we can show that $U_n'(\theta) - U_n''(\theta)$ is stochastically equicontinuous. Hence, $U_n'(\theta) - U_n''(\theta) \xrightarrow{\mathbb{P}} 0$ uniformly in θ. We then observe that

$$\begin{aligned} U_n(\theta) - U_n'(\theta) &= \widetilde{U}_{1,n}(\theta) + \widetilde{U}_{2,n}(\theta) + \widetilde{U}_{3,n}(\theta) \\ \widetilde{U}_{1,n}(\theta) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\left(\hat{\varepsilon}_{j,i\Delta_n} - \varepsilon_{j,i\Delta_n}\right)\varepsilon_{l,i\Delta_n} \\ \widetilde{U}_{2,n}(\theta) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\,\varepsilon_{j,i\Delta_n}\left(\hat{\varepsilon}_{l,i\Delta_n} - \varepsilon_{l,i\Delta_n}\right) \\ \widetilde{U}_{3,n}(\theta) &= \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\left(\hat{\varepsilon}_{j,i\Delta_n} - \varepsilon_{j,i\Delta_n}\right)\left(\hat{\varepsilon}_{l,i\Delta_n} - \varepsilon_{l,i\Delta_n}\right). \end{aligned}$$

We denote $f_{j,s}(\theta) = f_j(Z_s, V_s; \theta)$ and $\hat{f}_{j,i}(\theta) = f_j(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n}; \theta)$. Then $\hat{\varepsilon}_{j,i\Delta_n} - \varepsilon_{j,i\Delta_n} = f_{j,i\Delta_n}(\theta_0) - \hat{f}_{j,i}(\theta)$. By Lemma 2,

$$\begin{aligned} & \widetilde{U}_{3,n}(\theta) \\ & = \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{d}_{j,i}(\theta)\,\hat{d}_{l,i}(\theta)\left(f_{j,i\Delta_n}(\theta_0) - \hat{f}_{j,i}(\theta)\right)\left(f_{l,i\Delta_n}(\theta_0) - \hat{f}_{l,i}(\theta)\right) \\ & \xrightarrow{\mathbb{P}} \int_0^T d_j(Z_s, V_s; \theta)\, d_l(Z_s, V_s; \theta)\left(f_{j,s}(\theta_0) - f_{j,s}(\theta)\right)\left(f_{l,s}(\theta_0) - f_{l,s}(\theta)\right), \end{aligned}$$

where the convergence holds uniformly in θ. By Lemma 2, we also have $\widetilde{U}_{1,n}(\cdot) \xrightarrow{\mathbb{P}} 0$ and $\widetilde{U}_{2,n}(\cdot) \xrightarrow{\mathbb{P}} 0$ uniformly in θ. Hence, $U_n(\theta) - U_n'(\theta) \xrightarrow{\mathbb{P}} 0$ uniformly. Hence, $U_n(\theta) - U_n'(\theta)$, $U_n'(\theta) - U_n''(\theta)$ and $U_n''(\theta) - U(\theta)$ are all $o_p(1)$ uniformly in θ. It readily follows that $U_n(\theta) - U(\theta) = o_{pu}(1)$. This finishes the proof of (7.12). The assertions of the theorem readily follow. □

7.5 Proofs in Section 2.5

Proof of Theorem 4. By Lemma 5, $\widehat{G}_n(\cdot) \xrightarrow{\mathbb{P}} G(\cdot)$ uniformly on Θ. By Assumption H2,

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\widehat{B}_n(\theta)\right\| &\le \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \sup_{\theta\in\Theta}\left\|\gamma\left(Z_{i\Delta_n}, \widehat{V}_{i\Delta_n};\theta\right)\right\| \\ &\le K\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n}\left(1+\left|\widehat{V}_{i\Delta_n}\right|^{2p-2}\right)\widehat{V}^2_{i\Delta_n} \\ &= O_p(1),\end{aligned}$$

where the last line is due to Lemma 2. Hence, $G^\star_n(\cdot) \xrightarrow{\mathbb{P}} G(\cdot)$ uniformly on Θ. Following a similar argument as in Theorem 1, $\theta^\star_n \xrightarrow{\mathbb{P}} \theta_0$.

Let $H^\star_n(\theta) = \partial_\theta G^\star_n(\theta)$. Routine manipulation yields, w.p.a.1, $\theta^\star_n - \theta_0 = M^\star_n G^\star_n(\theta_0)$, where $M^\star_n = -(H^\star_n(\theta^\star_n)^\intercal \Xi^\star_n H^\star_n(\bar{\theta}^\star_n))^{-1} H^\star_n(\theta^\star_n)^\intercal \Xi^\star_n$ and $\bar{\theta}^\star_n$ is some mean value between $\theta^\star_n$ and θ_0. Since $H^\star_n(\theta) = \widehat{H}_n(\theta) + k_n^{-1}\partial_\theta \widehat{B}_n(\theta)$, by Proposition 1(a) and (7.7), $H^\star_n(\cdot) \xrightarrow{\mathbb{P}} H(\cdot)$ uniformly in a neighborhood containg θ_0. Hence, $H^\star_n(\theta^\star_n) \xrightarrow{\mathbb{P}} H$ and $H^\star_n(\bar{\theta}^\star_n) \xrightarrow{\mathbb{P}} H$, yielding $M^\star_n \xrightarrow{\mathbb{P}} M$. By Proposition 5, $G^\star_n(\theta_0) = \Delta_n^{1/2}\xi_n + o_p(\Delta_n^{1/2})$. The assertion then readily follows. □

Proof of Theorem 5. To simply notations, we denote $\Xi_n(\theta) = \widehat{S}_n(\theta)^{-1}$ and $\Xi(\theta) = S(\theta)^{-1}$. As shown in the proof of Theorem 4, $G^\star_n(\theta) - G(\theta) = o_{pu}(1)$. By (7.12), $\widehat{S}_n(\theta) - S(\theta) = o_{pu}(1)$, yielding $\Xi_n(\theta) - \Xi(\theta) = o_{pu}(1)$ because $|\det(S(\theta))|^{-1}$ is uniformly bounded by assumption. Hence, $Q^c_n(\theta) \xrightarrow{\mathbb{P}} Q^c(\theta) = G(\theta)^\intercal \Xi(\theta) G(\theta)$ uniformly in $\theta \in \Theta$. By bounded convergence, $Q^c(\cdot)$ is continuous. It is clear that θ_0 is the unique minimizer of $Q^c(\cdot)$, by Assumption GMM1(ii) and the assumption that $S(\theta)$ is positive definite. By a standard argument (e.g. Theorem 2.1, Newey and McFadden (1994)), we deduce $\theta^c_n \xrightarrow{\mathbb{P}} \theta_0$.

The first order condition (see (3.103) in Hall (2005)) and a mean-value expansion yield

$$\begin{aligned}0 &= \partial_\theta G^\star_n(\theta^c_n)^\intercal \Xi_n(\theta^c_n) G^\star_n(\theta^c_n) + \Lambda_n \mathcal{G}_n(\theta^c_n) \\ &= \partial_\theta G^\star_n(\theta^c_n)^\intercal \Xi_n(\theta^c_n) G^\star_n(\theta_0) + \partial_\theta G^\star_n(\theta^c_n)^\intercal \Xi_n(\theta^c_n)\partial_\theta G^\star_n(\bar{\theta}_n)(\theta^c_n - \theta_0) \\ &\quad + \Lambda_n \mathcal{G}_n(\theta_0) + \Lambda_n \partial_\theta \mathcal{G}_n\left(\bar{\theta}_n\right)(\theta^c_n - \theta_0),\end{aligned}$$

where $\Lambda_n \equiv -(\partial_\theta vec(\widehat{S}_n(\theta^c_n)))^\intercal (\Xi_n(\theta^c_n) \otimes \Xi_n(\theta^c_n))/2$, $\mathcal{G}_n(\theta) = vec(G^\star_n(\theta) G^\star_n(\theta)^\intercal)$ and $\bar{\theta}_n$ is some mean value betweeen θ^c_n and θ_0. By results in the proof of Theorem 4,

$$\partial_\theta G^\star_n(\theta^c_n) \xrightarrow{\mathbb{P}} H, \quad \partial_\theta G^\star_n\left(\bar{\theta}_n\right) \xrightarrow{\mathbb{P}} H, \quad G^\star_n(\bar{\theta}_n) \xrightarrow{\mathbb{P}} 0.$$

Hence, $\partial_\theta \mathcal{G}_n(\bar{\theta}_n) = o_p(1)$. By Lemma 2 and $\theta^c_n \xrightarrow{\mathbb{P}} \theta_0$, we derive $\partial_\theta \widehat{S}_n(\theta) = O_p(1)$. Hence,

$\Lambda_n = O_p(1)$. By Propositions 5 and 6, $\mathcal{G}_n(\theta_0) = O_p(\Delta_n)$. Therefore, w.p.a.1,

$$\begin{aligned}\theta_n^c - \theta_0 &= -\left(\partial_\theta G_n^\star(\theta_n^c)^\intercal \Xi_n(\theta_n^c) G_n^\star(\theta_0) + \Lambda_n \partial_\theta \mathcal{G}_n\left(\bar{\theta}_n\right)\right)^{-1} \\ &\quad \times \left(\partial_\theta G_n^\star(\theta_n^c)^\intercal \Xi_n(\theta_n^c) G_n^\star(\theta_0) + \Lambda_n \mathcal{G}_n\left(\theta_0\right)\right) \\ &= -\left(H^\intercal S^{-1} H + o_p\left(1\right)\right)^{-1} \left(H^\intercal S^{-1} \Delta_n^{1/2} \xi_n + o_p(\Delta_n^{1/2})\right) \\ &= -\Delta_n^{1/2} \left(H^\intercal S^{-1} H\right)^{-1} H^\intercal S^{-1} \xi_n + o_p(\Delta_n^{1/2}).\end{aligned}$$

□

7.6 Proofs in Section 3

Proof of Theorem 6. Below, let ϑ_n be θ_n^* or $\theta_n^\star$. By Theorems 2 and 4,

$$\Delta_n^{-1/2}(\vartheta_n - \theta_0) = M\xi_n + o_p\left(1\right), \quad \widehat{B}_n\left(\theta_0\right) - \widehat{B}_n\left(\vartheta_n\right) = O_p(\Delta_n^{1/2}). \tag{7.13}$$

where the second line follows a mean-value expansion and (7.7). Hence,

$$\begin{aligned}G_n^\star(\vartheta_n) &= \widehat{G}_n\left(\theta_0\right) + \widehat{H}_n\left(\bar{\theta}_n\right)\left(\vartheta_n - \theta_0\right) - k_n^{-1}\widehat{B}_n\left(\vartheta_n\right) \\ &= \Delta_n^{1/2}\xi_n + H\left(\vartheta_n - \theta_0\right) + k_n^{-1}\left(\widehat{B}_n\left(\theta_0\right) - \widehat{B}_n\left(\vartheta_n\right)\right) + o_p\left(\Delta_n^{1/2}\right) \\ &= \Delta_n^{1/2}(I_k + HM)\xi_n + o_p\left(\Delta_n^{1/2}\right),\end{aligned}$$

where the first equality is by a mean-value expansion with mean-value $\bar{\theta}_n$ between θ_0 and ϑ_n; the second equality follows Proposition 1(a) and Proposition 5, the third equality is due to (7.13). By Proposition 6, ξ_n converges stably in law to a mixture centered Gaussian variable with $\mathcal{F}$-conditional variance-covariance matrix S. The assertion of the theorem then follows a standard argument for the J-statistic. □

Proof of Theorem 7. The first assertion follows a similar argument as in the proof of Theorem 2. The asymptotic $\mathcal{F}$-conditonal independence between $\theta_n^*\left(\tau\right)$ and $\theta_n^*\left(\tau'\right)$ for $\tau \neq \tau'$ follows the approximate martingale property of α_i^n and $\varepsilon_{i\Delta_n}$. The second assertion follows the continuous mapping theorem and the first assertion. □

Proof of Corollary 1. (a) The assertion follows Theorem 7 by the continuous mapping theorem. (b) Note that $\widehat{\mathbf{\Sigma}}_n \xrightarrow{\mathbb{P}} \mathbf{\Sigma}$. Upon using a subsequence argument, we can assume that $\widehat{\mathbf{\Sigma}}_n \to \mathbf{\Sigma}$ a.s.. Hence, conditionally on $\mathcal{F}$, $\mathcal{S}((R^\intercal \widehat{\mathbf{\Sigma}}_n R)^{1/2}\tilde{\zeta}; \widehat{\mathbf{\Sigma}}_n)$ converges in distribution to the $\mathcal{F}$-conditional law of $\mathcal{S}(\tilde{\xi}; \mathbf{\Sigma})$. The assertion really follows as the quantile function of $\mathcal{S}(\tilde{\xi}; \mathbf{\Sigma})$ is continuous at $1-\alpha$ by the condition in part (b). (c) The assertion follows Theorem 7 and property (iii) of $\mathcal{S}$. □

Proof of Corollary 2. The assertion is readily implied by Theorem 6. $\square$

Proof of Proposition 2. By Proposition 5, $G_n^\star(\theta_0) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, S)$. We also note that $\widehat{S}_n(\theta_0) \xrightarrow{\mathbb{P}} S$. The assertion then follows the continuous mapping theorem. $\square$

7.7 The VIX Calculation

Proof of Proposition 3. Note that we have for any $s \geq t$,

$$dE_t(V_s) = \kappa(\xi - E_t(V_s))ds,$$

so that

$$E_t\Big(\frac{1}{\tau}\int_t^{t+\tau} V_s ds\Big) = \xi + \frac{1 - e^{-\kappa\tau}}{\kappa\tau}(V_t - \xi).$$

Therefore,

$$\begin{aligned} y_t &= \frac{1}{\tau}E_t\Big(\int_t^{t+\tau} V_s ds + \int_t^{t+\tau}\int_{\mathbb{R}} j(x, V_s)^2 \nu(V_s, dx)ds\Big) \\ &= \gamma + \xi(1+\eta)\Big(1 - \frac{1 - e^{-\kappa\tau}}{\kappa\tau}\Big) + (1+\eta)\frac{1 - e^{-\kappa\tau}}{\kappa\tau}V_t, \end{aligned}$$

So we have

$$\theta_1 = \gamma + \xi(1+\eta)\Big(1 - \frac{1 - e^{-\kappa\tau}}{\kappa\tau}\Big) \quad \text{and} \quad \theta_2 = (1+\eta)\frac{1 - e^{-\kappa\tau}}{\kappa\tau}.$$

This concludes the proof. $\square$.

Proof of Proposition 4. Let $\psi(u)$ be the cumulant function of the underlying Lévy process L_t. That is,

$$E\exp(uL_t) = \exp(t\psi(u)).$$

We assume for now that for all $u \geq 0$,

$$E\exp(uL_t) < \infty.$$

By Lévy-Khinchin representation, we have

$$\psi(u) = \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}}(e^{uf} - 1 - uf)\mu(df).$$

Note that we can write

$$F_t = e^{\kappa t}\Big(F_0 + \int_0^t e^{-\kappa s} dL_s\Big),$$

where we have

$$L_t = \sigma B_t^Q + \int_0^t \int_{\mathbb{R}} f\Big(M(ds, df) - \mu(df)ds\Big)\Big),$$

a Lévy martingale. Moreover, it has been shown in Novikov (2003) that, if

$$E \log\Big(1 + \max(-L_1, 0)\Big) < \infty,$$

then the exponentiated process

$$\mathcal{E}_t(u) = \exp\Big(ue^{-\kappa t}F_t - \int_0^t \psi(ue^{\beta v})dv\Big)$$

is a martingale. This implies that

$$E_t \exp(\beta F_s) = \exp\Big(\beta e^{\kappa(s-t)}F_t + \int_t^s \psi(\beta e^{\kappa(s-v)})dv\Big).$$

Therefore, we have

$$\begin{aligned} E_t(V_s) &= E_t \exp(\alpha + \beta F_s) \\ &= \exp\Big(\alpha + \beta e^{\kappa(s-t)}F_t + \int_t^s \psi(\beta e^{\kappa(s-v)})dv\Big) \\ &= \exp\Big(\alpha + e^{\kappa(s-t)}(\log V_t - \alpha) + \int_t^s \psi(\beta e^{\kappa(s-v)})dv\Big). \end{aligned}$$

Thus, we have

$$(\text{VIX}/100)_t^2 = \frac{1}{\tau}E_t\Big(\int_t^{t+\tau} \gamma + (\eta+1)V_s ds\Big)$$

$$= \frac{1}{\tau}\int_0^\tau \gamma + (\eta+1)\exp\Big(\alpha + e^{\kappa u}(\log V_t - \alpha) + \int_0^u \psi(\beta e^{\kappa v})dv\Big)du.$$

Note that we only need $\psi(u) < \infty$, for $u \in [\beta, \beta e^{\kappa\tau}]$ or $[\beta e^{\kappa\tau}, \beta]$. □