Spot Variance Regressions *

Jia Li† Dacheng Xiu‡
Duke University Chicago Booth

This Version: February 10, 2013

Abstract

We study a nonlinear vector regression model for discretely sampled high-frequency data with the latent spot variance of an asset as a covariate. 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 the standard inference implausible. We propose several bias-correction methods and show that the bias-corrected estimators have the parametric rate of convergence with mixture normal asymptotic distributions. We provide estimators for the asymptotic variance, as well as Anderson-Rubin-type confidence sets for the true parameter. Tests for overidentification and parameter stability are constructed. An empirical application on VIX pricing provides substantive evidence against conventional risk-neutral models for volatility dynamics.

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

JEL Codes: C22.

*We are grateful to Yacine Aït-Sahalia, Alan Bester, Tim Bollerslev, Federico Bugni, Chris Hansen, Oliver Linton, Per Mykland, Andrew Patton, Jeff Russell, George Tauchen, Viktor Todorov, Lan Zhang, as well as conference and seminar participants at the University of Chicago and the 2012 Triangle Econometrics Conference for their helpful comments. Li’s work was partially supported by NSF Grant SES-1227448. Xiu’s work was supported in part by the FMC Faculty Scholar Fund at the University of Chicago Booth School of Business.

†Box 90097, Duke University, Durham, NC, 27708. Email: jl410@duke.edu.

‡University of Chicago Booth School of Business, 5807 S. Woodlawn Avenue, Chicago, IL 60637. Email: dacheng.xiu@chicagobooth.edu.
1 Introduction

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

\[ Y_t = f (Z_t, V_t; \theta) + \varepsilon_t, \]  

(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 \( \theta \), \( X_t \) 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 \( \varepsilon_t \) is the pricing error.

Direct 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

\[ X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_t dW_t + J_t, \]  

(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 convoluted with not only 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 to conduct joint inference. Such a 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 the 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 “raw” 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 is constructed from (possibly discontinuous) semimartingales that
are unique to the analysis of high-frequency data.

We show that the raw estimator is consistent under the fill-in asymptotics with a fixed time span. A striking result of this paper is that the raw estimator, when centered at the true parameter value, does not achieve a central limit theorem (CLT). Instead, the second-order asymptotic behavior of the raw 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 to the standard two-stage GMM setting, where a CLT is available and the first-stage estimation only affects the asymptotic variance (cf. Newey and McFadden (1994)).

We characterize the higher-order bias in the raw estimator and propose a simple bias-correction. We show that the bias-corrected estimator is centered at the true parameter value and admits a CLT with a parametric rate of convergence. Consistent estimators for the asymptotic variance are given in explicit forms for the purpose of inference.

We further show that some classical results in the GMM literature can be adapted to the current nonstandard setting. First, we propose a continuous updating estimator in the sense of Hansen, Heaton, and Yaron (1996), which achieves both bias-correction and efficient choice of weighting matrix within one step. Second, we construct Anderson-Rubin-type confidence sets for the true parameter similarly as in Stock and Wright (2000) and Andrews and Soares (2010). Finally, we construct overidentification tests for misspecification and inference tools in order to examine parameter stability.

Empirically, we apply the method to a linear pricing model for the VIX, where the linear specification is implied by a large class of structural models, including the popular affine jump-diffusion model as a special case. We document strong evidence for parameter instability. We further generalize the linear model to a nonlinear model, which we label as the “ABC” model. We show that the ABC model serves as a parsimonious approximation for a structural VIX pricing model based on exponential-Lévy risk-neutral volatility dynamics. When nested in the ABC model, the linear specification is strongly rejected.

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. This 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 the leverage effect and volatility jumps with infinite activity or even infinite variation. This 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 rich information of such datasets. We demonstrate the usefulness of these tools with a novel empirical application on VIX pricing.

Turning to the related literature, the current paper is closely related to the literature on nonparametric inference for volatility functionals; see Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Gonçalves and Meddahi (2009), Todorov and Tauchen (2012), Li, Todorov, and Tauchen (2012) and, in particular, the recent work of Jacod and Rosenbaum (2012). However, 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. Andersen, Fusari, and Todorov (2012) infer latent state variables from (asymptotically large) daily option panels and use only the spot variance estimate at the market close to construct specification tests. Our approach does not require the availability of large option panels; instead, it makes full use of high-frequency data, leading to theoretical results distinct from those in Andersen, Fusari, and Todorov (2012).

The paper is clearly related to the vast literature on GMM, in particular Hansen (1982) and Hansen, Heaton, and Yaron (1996). The Anderson-Rubin-type confidence sets proposed in Section 2.6 are similar to those of Stock and Wright (2000) and Andrews and Soares (2010). That said, our theory is very different from the existing literature due to the complications of the first-stage nonparametric recovery of the stochastic volatility path from the price returns, hence 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 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. Section 4 shows simulation results, followed by an empirical application in Section 5. Section 6 concludes. All proofs are in the online supplement to this paper.

---

1For contributions and extensions on GMM in the empirical option pricing literature, see Pan (2002), Gagliardini, Gouriéroux, and Renault (2011) and references therein.
All limits below are taken as \( n \to \infty \). All vectors are column vectors. The transpose of any matrix \( A \) is denoted by \( A^\top \). For notational simplicity, we write \((a, b)\) instead of \((a^\top, b^\top)^\top\). 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 \( \partial_x f(x, y)^\top \) its \( \dim(x) \times q \) partial derivative matrix w.r.t. \( x \), with the transpose sign suppressed when \( q = 1 \). Similarly, \( \partial^2_{xx} f(x, y) \) denotes the \( \dim(x) \times \dim(x) \) Hessian matrix of \( f(x, y) \) w.r.t. \( x \) and \( \partial_{x,y}^2 f(x, y) \) is the \( j \)th 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 \( 1 \leq j \leq J \),

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

(2.1)

where \( Y_{j,t}, Z_t \) and \( V_t \) take values in \( Y \subseteq \mathbb{R}, Z \subseteq \mathbb{R}^{\dim(Z)} \) and \( V \subseteq \mathbb{R}^+ \) respectively, \( \theta_0 \in \Theta \subseteq \mathbb{R}^{\dim(\Theta)} \) is the true parameter and the parameter space \( \Theta \) is compact. The latent spot variance \( V_t \) is related 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 \( \Delta_n^i X = X_{i\Delta_n} - X_{(i-1)\Delta_n} \).

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

\[
\hat{V}_{i\Delta_n} \equiv \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} (\Delta_n^{i+j} X)^2 1_{\{ |\Delta_n^{i+j} X| \leq u_n \}},
\]

(2.2)

where the truncation threshold \( u_n \equiv \bar{\alpha} \Delta_n^\varpi \) for constants \( \bar{\alpha} > 0 \) and \( \varpi \in (0, 1/2) \). The truncation indicator in (2.2) 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

\[
\hat{G}_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} g(Y_{i\Delta_n}, Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta).
\]
For a sequence $\Xi_n$ of weighting matrices converging in probability to a $\mathcal{F}$-measurable random matrix $\Xi$, we define the GMM estimator as follows:

$$\hat{\theta}_n \equiv \arg \min_{\theta \in \Theta} \hat{Q}_n(\theta), \quad \text{where} \quad \hat{Q}_n(\theta) \equiv \hat{G}_n(\theta)^T \Xi_n \hat{G}_n(\theta). \quad (2.3)$$

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

$$\hat{G}_n(\theta) \xrightarrow{p} G(\theta), \quad \hat{Q}_n(\theta) \xrightarrow{p} Q(\theta) \quad (2.4)$$

where

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

Notice that the GMM problem considered here is 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) $\Theta$ is compact; (b) $\theta_0 \in \text{int}(\Theta)$ is the unique solution to $G(\theta) = 0$ a.s.; (c) $\Xi_n \xrightarrow{p} \Xi$ for some $\mathcal{F}$-measurable $k \times k$ matrix $\Xi$, which is a.s. finite and positive definite.

Assumption GMM1 is mainly concerned with the identification of the true parameter $\theta_0$. The key assumption is the uniqueness of $\theta_0$ as a solution to $G(\theta) = 0$ a.s., which, combined with the positive definiteness of $\Xi$, ensures that $\theta_0$ 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) $Y^*$, $Z$ and $X$ are càdlàg adapted processes 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 $\sigma_t$ càdlàg. Moreover, $|\delta(\omega, t \land \tau_m(\omega), z)| \wedge 1 \leq \Gamma_m(z)$ for all $\omega, t, z$, where $(\tau_m)$ is a localizing sequence of stopping times and, for some $r \in (0, 1)$, each function $\Gamma_m$ on $\mathbb{R}$ satisfies $\int_{\mathbb{R}} \Gamma_m(z)^r \lambda(dz) < \infty$. 

6
(b) The process $\sigma_t$ is also an Itô semimartingale with the form

$$
\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s \, ds + \int_0^t \tilde{\sigma}_s \, dW_s + \int_0^t \tilde{\sigma}_s' \, dW'_s
$$

$$
+ \int_0^t \int_\mathbb{R} \tilde{\delta}(s, z) \mathbf{1}_{\{\tilde{\delta}(s, z) \leq 1\}} (\mu - \nu) \, (ds, dz)
$$

$$
+ \int_0^t \int_\mathbb{R} \tilde{\delta}(s, z) \mathbf{1}_{\{\tilde{\delta}(s, z) > 1\}} \mu \, (ds, dz),
$$

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

Assumption H1 is fairly unrestrictive and satisfied by most continuous-time models in finance. We allow price jumps with 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 not completely general. For example, we restrict the price jumps to have finite variation, but this is a standard condition (see Jacod and Protter (2012)) for ensuring that the truncation-based spot variance estimator is robust to jumps in the analysis of second-order asymptotics.

**Assumption H2.** For some $p \geq 3$, we have $\omega \in [(4p - 1)/2(4p - r), 1/2)$, where $r$ is defined in Assumption H1(a). Moreover, for any compact $K_2 \subseteq \mathcal{Z}$, there exists some constant $C > 0$ such that for all $v \in \mathcal{V}$,

(a) for each $\theta \in \Theta$, $\sup_{z \in K_2} (\|\partial_v^m d_j(z, v; \theta)\| + \|\partial_v^m f_j(z, v; \theta)\|) \leq C(1 + |v|^{p-m})$ for $m = 0, 1, 2, 3$;

(b) $\sup_{\theta \in \Theta, z \in K_2} (\|\partial_v^m d_j(z, v; \theta)\| + \|\partial_v^m f_j(z, v; \theta)\|) \leq C(1 + |v|^p)$, $m = 1, 2$;

(c) $\sup_{\theta \in \Theta, z \in K_2} (\|\partial_\theta \partial_v d_j(z, v; \theta)\| \leq C(1 + |v|^{p-1})$;

(d) $\sup_{\theta \in \Theta, z \in K_2} (\|\partial_\theta \partial_v^m f_j(z, v; \theta)\| \leq C(1 + |v|^{p-m})$, $m = 1, 2$;

(e) all partial derivatives in (a)-(d) are continuous in $(z, v, \theta)$.

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.\footnote{The polynomial growth condition is typically used in the study of spot volatility estimators, see e.g. Jacod and Protter (2012) and Jacod and Rosenbaum (2012). This condition is fairly unrestrictive near infinity, i.e. when volatility is large, but does not accommodate functions which are explosive at zero, e.g. the logarithm function. For functions that are ill-behaved near zero, one may instead consider “regularized” functional forms in order to fulfill this technical requirement. For example, the logarithm function may be replaced by a smooth function which coincides with the logarithm function on $(\underline{\omega}, \infty)$ for some small $\underline{\omega} > 0$ and is bounded with bounded derivatives on $[0, \underline{\omega}]$. While not completely theoretically satisfactory, the standard polynomial growth condition is maintained and we recommend the aforementioned regularization as a reasonable practical compromise. After all, scenarios in which the volatility is very close to zero is unlikely to be economically interesting. The technical maneuver for relaxing such conditions is left for future research.} This condition is needed to tame the
effect of the estimation error of the spot variance in the second-stage GMM estimation. For the purpose of proving uniform convergence, the conditions are stated uniformly over $\Theta$ and $K Z$, but they are not overly restrictive because $\Theta$ and $K Z$ are compact, so that boundedness is implied by smoothness conditions.

**Assumption H3.** Conditionally on $F$, 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 $t \mapsto \mathbb{E} \left[ \|\varepsilon_t\|^4 | F \right]$ is $F_t$-adapted and locally bounded. Moreover, $A_t \equiv \mathbb{E} [\varepsilon_t \varepsilon_t^\top | F]$ is $F_t$-adapted and càdlàg.

Assumption H3 is akin to the assumption on microstructure noise employed in the noise-robust estimation of volatility and jump functionals; see Jacod, Li, Mykland, Podolskij, and Vetter (2009), Jacod, Podolskij, and Vetter (2010), Aït-Sahalia, Jacod, and Li (2012) and Li (2012) for details. This connection is not surprising since (2.1) imposes a semimartingale-plus-noise structure for the dependent variable. The somewhat strong requirement is the conditional independence of $\varepsilon_t$ as it rules out the autocorrelation of the pricing error (Hansen and Lunde (2006)). That said, Assumption H3 does allow for heteroskedasticity in $\varepsilon_t$, unconditional dependence between $\varepsilon_t$ and the underlying processes as well as serial dependence in the $\varepsilon_t$ process itself through higher moments.

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

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

To simplify the exposition, we maintain Assumptions GMM1, H1-H4, as well as GMM2 and GMM3 below throughout the paper without further mention. Under (a subset of) these assumptions, $\hat{\theta}_n$ is a consistent estimator for $\theta_0$, as shown below.

**Theorem 1** (a) We have (2.4) uniformly in $\theta \in \Theta$; (b) $\hat{\theta}_n \xrightarrow{P} \theta_0$.

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

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

$$\hat{\theta}_n - \theta_0 = - (\tilde{H}_n(\hat{\theta}_n)^\top \Xi_n \tilde{H}_n(\hat{\theta}_n))^{-1} \tilde{H}_n(\hat{\theta}_n)^\top \Xi_n \tilde{G}_n(\theta_0),$$  

(2.5)

\footnote{See Newey and McFadden (1994) and the proof of Proposition 1 for details.}
where \( \tilde{H}_n (\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]} - k_n \nabla \theta g(Y_{i\Delta_n}, Z_{i\Delta_n}, \tilde{v}_{i\Delta_n}; \theta) \) is the \( k \times \text{dim} (\theta) \) Jacobian matrix associated with the sample moment function \( \tilde{G}_n (\theta) \) and \( \hat{\theta}_n \) is a mean value on the line joining \( \theta_0 \) and \( \hat{\theta}_n \) which may change from element to element in the vector equation. To describe the asymptotic behaviors of \( \hat{H}_n (\cdot) \) and \( \tilde{G}_n (\theta_0) \), we set

\[
\begin{align*}
H (\theta) &\equiv \int_0^T \partial \theta g (Y_s^*, Z_s, V_s; \theta) \, ds, \\
B &\equiv \int_0^T \partial^2 \theta g (Y_s^*, Z_s, V_s; \theta_0) V_s^2 \, ds.
\end{align*}
\]

(2.6)

**Assumption GMM2.** \( \hat{H}' \Xi \hat{H} \) is a.s. nonsingular.

**Proposition 1** Let \( \Theta_0 \) be some neighborhood containing \( \theta_0 \). We have (a) \( \sup_{\theta \in \Theta_0} \| \hat{H}_n (\theta) - H (\theta) \| = o_p (1) \); (b) \( k_n \tilde{G}_n (\theta_0) \xrightarrow{p} B \); (c) \( k_n (\hat{\theta}_n - \theta_0) \xrightarrow{p} -(\hat{H}' \Xi \hat{H})^{-1} \hat{H}' \Xi B \).

Proposition 1 illustrates the standard, as well as the non-standard, aspects of the asymptotic behavior of \( \hat{\theta}_n \). Part (a) is analogous to the convergence of the Jacobian matrix in a standard GMM setting. Although \( \tilde{V}_{i\Delta_n} \) 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. However, part (b) shows that the scaled sample moment \( k_n \tilde{G}_n (\theta_0) \) does not fulfill a CLT as in the standard setting; 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 \( \tilde{V}_{i\Delta_n} \), 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 raw estimator \( \hat{\theta}_n \). The lack of CLT for \( \hat{\theta}_n \) renders standard GMM inference procedure implausible. Importantly, it invalidates the “naive” approach of treating \( \tilde{V}_{i\Delta_n} \) as if it were \( V_{i\Delta_n} \) without error. That said, the positive message of Proposition 1(c) is that \( \hat{\theta}_n \) is a \( k_n \)-consistent estimator for \( \theta_0 \)—an improvement relative to the consistency result in Theorem 1.

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 setting, the instruments are \( (1, v) \), so we have \( g (y, v; \theta) = (\alpha + \beta v - y, v (\alpha + \beta v - y)) \).

\(^{4}\)This higher-order bias does not arise because of the undersmoothing condition \( k_n^2 \Delta_n \rightarrow 0 \). With the “optimal” choice of \( k_n \), i.e. \( k_n \asymp \Delta_n^{-1/2} \), Jacod and Rosenbaum (2012) show that this bias term is still present, but further confounded by two other sources of bias due to volatility of volatility and volatility jumps; see Theorem 3.1 there. The undersmoothing condition only makes the higher-order bias term easier to characterize and correct, so that feasible inference can be conducted.
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

$$k_n \left( \hat{\alpha}_n - \alpha_0 \right) \xrightarrow{p} -H^{-1} B = \frac{2\beta_0}{1 - IV^2/IQ} \begin{pmatrix} IV \\ IV^2/IQ \end{pmatrix},$$

where $H = \begin{pmatrix} 1 & IV \\ IV & IQ \end{pmatrix}$, $B = \begin{pmatrix} 0 \\ 2\beta_0 IQ \end{pmatrix}$.

By the Cauchy-Schwarz inequality, $IQ > IV^2$ as soon as $V_t$ is not constant over $[0, T]$. The relative bias (normalized by $\beta_0$) 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., when $V_t$ has low variation over $[0, T]$.

### 2.3 Bias correction

Proposition 1 suggests that the second-order bias in $\hat{\theta}_n$ arises from the moment function $\hat{G}_n(\theta)$. A natural approach for bias-correction is to estimate the bias term $B$, and then correct for it in the moment function. By direct calculation, we can rewrite $B = \int_0^T \gamma(Z_s, V_s; \theta_0) ds$, where $\gamma(z, v; \theta) = (\gamma_1(z, v; \theta), \ldots, \gamma_J(z, v; \theta))$ and for $1 \leq j \leq J$,

$$\gamma_j(z, v; \theta) = (2\partial_\nu d_j(z, v; \theta) \partial_v f_j(z, v; \theta) + d_j(z, v; \theta) \partial^2_v f_j(z, v; \theta)) v^2.$$

We consider the following sample analogue for $\int_0^T \gamma(Z_s, V_s; \theta) ds$:

$$\hat{B}_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \gamma(Z_i\Delta_n, \hat{V}_{i\Delta_n}; \theta).$$

The bias-corrected moment function is $G_\star_n(\theta) \equiv \hat{G}_n(\theta) - k_n^{-1} \hat{B}_n(\theta)$, with which we associate the objective function $Q_\star_n(\theta) = G_\star_n(\theta)^\top \Xi_n G_\star_n(\theta)$. The bias-corrected estimator is then defined as

$$\theta_\star_n = \arg\min_{\theta \in \Theta} Q_\star_n(\theta). (2.7)$$

We refer to $\theta_\star_n$ as the one-step estimator (cf. Section 2.7).

**Linear Regression Example—Continued.** Let $\ell_n = \Delta_n([1/\Delta_n] - k_n + 1)$, $IV_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \hat{V}_{i\Delta_n}$, $IQ_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \hat{V}_{i\Delta_n}^2$, $\nabla_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} Y_{i\Delta_n}$ and $VY_n = \Delta_n \sum_{i=0}^{[1/\Delta_n]-k_n} \hat{V}_{i\Delta_n} Y_{i\Delta_n}$. The raw estimator and the one-step (bias-corrected) estimator are re-
respectively given by
\[
\hat{\theta}_n = \left( \begin{array}{cc} \iota_n & IV_n \\ IV_n & IQ_n \end{array} \right)^{-1} \left( \begin{array}{c} \Upsilon_n \\ VY_n \end{array} \right),
\]
\[
\theta^*_n = \left( \begin{array}{cc} \iota_n & IV_n \\ IV_n & (1 - \frac{2}{k_n}) IQ_n \end{array} \right)^{-1} \left( \begin{array}{c} \Upsilon_n \\ VY_n \end{array} \right).
\]

Theorem 2 below shows that \( \theta^*_n \) has a well-behaved asymptotic distribution. 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
\[
(j,l) \text{ block of } S_1(\theta) \equiv \int_0^T d_j(Z_s,V_s;\theta) d_l(Z_s,V_s;\theta)^T A_{jl,s} ds,
\]
where \( A_{jl,s} \) is the \((j,l)\) component of \( A_s \) defined in Assumption H3. We also set \( \phi_j(z,v;\theta) \equiv d_j(z,v;\theta) \partial_v f_j(z,v;\theta), \) \( 1 \leq j \leq J \), and \( \phi(z,v;\theta) = (\phi_1(z,v;\theta), \ldots, \phi_J(z,v;\theta)) \) and consider the \( k \times k \) matrix
\[
S_2(\theta) \equiv 2 \int_0^T \phi(Z_s,V_s;\theta) \phi(Z_s,V_s;\theta)^T V_s^2 ds.
\]
Finally, let \( S(\theta) \equiv S_1(\theta) + S_2(\theta) \).

Below, for a sequence of random variables \( \chi_n \), the notation \( \chi_n \overset{L}{\rightarrow} \mathcal{MN}(0,\Sigma_\chi) \) indicates that \( \chi_n \) converges stably in law \(^5\) 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 \( \Sigma_\chi \).

Theorem 2 Let \( \theta^*_n \) be given by (2.7). We have \( \Delta_n^{-1/2}(\theta^*_n - \theta_0) = M\xi_n + o_p(1) \), where \( M \equiv -(H^\top \Xi H)^{-1} H^\top \Xi \) and the sequence of variables \((\xi_n)_{n \geq 1}\) satisfies \( \xi_n \overset{L}{\rightarrow} \mathcal{MN}(0,S(\theta_0)) \). Consequently, \( \Delta_n^{-1/2}(\theta^*_n - \theta_0) \overset{L}{\rightarrow} \mathcal{MN}(0,\Sigma) \), where \( \Sigma \equiv MS(\theta_0)M^\top \).

Comment. Similarly as in the standard GMM setting, the optimal choice of the weighting matrix is \( \Xi = S(\theta_0)^{-1} \), which minimizes the asymptotic variance \( \Sigma \) in the matrix sense under the usual partial order; the associated estimator \( \theta^*_n \) is referred to as the efficient one-step estimator.

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 \( \Sigma \). We first construct an estimator for \( S(\theta_0) \). Let \( \hat{\epsilon}_{j,i\Delta_n}(\theta) = Y_{j,i\Delta_n} - \)

\(^5\)Stable convergence is a slightly stronger notion than the usual weak convergence. This convergence concept is needed here because the asymptotic variance is random. Appendix A in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) provides a detailed discussion on stable convergence; also see Jacod and Shiryaev (2003) and Jacod and Protter (2012).
\[ f_j(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta). \] We define \( \widehat{S}_{1,n}(\theta) \) to be a \( k \times k \) matrix with its \((j,l)\) block, \( 1 \leq j, l \leq J \), given by
\[
(j, l) \text{ block of } \widehat{S}_{1,n}(\theta) \\
\equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} d_j(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta) d_l(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta) \psi_{j,i\Delta_n}(\theta) \psi_{l,i\Delta_n}(\theta)
\]
and set
\[
\widehat{S}_{2,n}(\theta) \equiv 2\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \phi(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta) \phi(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta) \psi_{i\Delta_n}^\top \psi_{i\Delta_n}.
\]
Finally, let \( \widehat{S}_n(\theta) \equiv \widehat{S}_{1,n}(\theta) + \widehat{S}_{2,n}(\theta) \).

**Theorem 3** Let \( \tilde{\theta}_n \) and \( \tilde{H}_n \) be preliminary estimators satisfying \( \tilde{\theta}_n = \theta_0 + o_p(1) \) and \( \tilde{H}_n = H + o_p(1) \). Let \( \tilde{M}_n \equiv (\tilde{H}_n \Xi \tilde{H}_n)^{-1} \tilde{H}_n \Xi_n \) and \( \hat{\Sigma}_n = \tilde{M}_n \hat{S}(\tilde{\theta}_n) \tilde{M}_n^\top \). Then \( \hat{S}(\tilde{\theta}_n) = S(\theta_0) + o_p(1) \) and \( \hat{\Sigma}_n = \Sigma + o_p(1) \).

**Comments.** (i) Theorem 3 shows that \( \hat{\Sigma}_n \) is a consistent estimator for \( \Sigma \). Confidence sets for \( \theta_0 \) or its subvector can be constructed in the usual way using the asymptotic mixture normal approximation in Theorem 2.

(ii) Theorem 3 also shows that \( \hat{S}(\tilde{\theta}_n) = S(\theta_0) + o_p(1) \) provided that \( \tilde{\theta}_n = \theta_0 + o_p(1) \). Hence, an efficient GMM procedure can be implemented by setting \( \Xi_n = \hat{S}(\tilde{\theta}_n)^{-1} \).

We can take the preliminary estimator \( \tilde{\theta}_n \) in Theorem 3 to be the raw estimator \( \hat{\theta}_n \) or the one-step estimator \( \theta^*_n \). There are many possible choices for \( \tilde{H}_n \). For example, for any \( \tilde{\theta}_n = \theta_0 + o_p(1) \), \( \tilde{H}_n(\theta_0) \) serves as a valid choice for \( \tilde{H}_n \) because of Proposition 1(a). Another choice of \( \tilde{H}_n \) is as follows. Let \( \psi_j(z,v; \theta) = d_j(z,v; \theta) \partial f_j(z,v; \theta) \psi_j(z,v; \theta)^\top, 1 \leq j \leq J \), and \( \psi(z,v; \theta) = (\psi_1(z,v; \theta), \ldots, \psi_J(z,v; \theta)) \). By direct calculation, \( \tilde{H} = \int_0^T \psi(Z_s, V_s; \theta_0) ds \). This representation motivates the following sample analogue estimator
\[
\tilde{H}_n(\theta) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \psi(Z_{i\Delta_n}, \hat{V}_{i\Delta_n}; \theta), \quad \theta \in \Theta.
\]

**Lemma 1** Suppose \( \hat{\theta}_n \xrightarrow{P} \theta_0 \). Then \( \tilde{H}_n(\hat{\theta}_n) \xrightarrow{P} H \).

### 2.5 Continuous updating estimator

The one-step estimator can be modified to a *continuous updating estimator* (Hansen, Heaton, and Yaron (1996)), which achieves the bias-correction and the optimal
choice of weighting matrix within the estimation step. The continuous updating estimator is given by

\[ \theta^c_n \equiv \arg\min_{\theta \in \Theta} Q^c_n(\theta), \quad Q^c_n(\theta) \equiv G_n^* (\theta)^\top \hat{S}_n(\theta)^{-1} G_n^* (\theta). \]

We need an additional assumption on \( S(\theta) \) to ensure that the random function \( Q^c_n(\cdot) \) is well-behaved asymptotically.

**Assumption GMM3.** \( \inf_{\theta \in \Theta} \lambda_{\min}(S(\theta)) > 0 \) a.s., where the function \( \lambda_{\min}(\cdot) \) computes the smallest eigenvalue of its argument.

The following theorem shows that the continuous updating estimator is asymptotically equivalent to the efficient one-step estimator.

**Theorem 4** We have \( \Delta_n^{-1/2}(\theta^c_n - \theta_0) \overset{d}{\to} \chi^2_k \) (chi-square distribution with degree of freedom \( k \));

(b) \( \operatorname{MaxS}_n(\theta_0) \) converges stably in law to \( \mathcal{M}(\hat{\xi}) \), where the \( k \times 1 \) variable \( \hat{\xi} \) is defined on an extension of the space \( (\Omega, \mathcal{F}, \mathbb{P}) \), which conditionally on \( \mathcal{F} \), is centered Gaussian with variance-covariance \( DS(\theta_0)^{-1/2} \hat{S}_n(\theta)^{-1/2} G_n^* (\theta) \), where for any \( k \)-vector \( x = (x_1, \ldots, x_k) \), \( \mathcal{M}(x) \equiv \max_{1 \leq j \leq k} |x_j| \). We first establish the asymptotic distributions of \( Q^c_n(\theta_0) \) and \( \operatorname{MaxS}_n(\theta_0) \).

**Proposition 2** (a) \( \Delta_n^{-1} Q^c_n(\theta_0) \overset{d}{\to} \chi^2_k \) (chi-square distribution with degree of freedom \( k \));

(b) \( \operatorname{MaxS}_n(\theta_0) \) converges stably in law to \( \mathcal{M}(\hat{\xi}) \), where the \( k \times 1 \) variable \( \hat{\xi} \) is defined on an extension of the space \( (\Omega, \mathcal{F}, \mathbb{P}) \), which conditionally on \( \mathcal{F} \), is centered Gaussian with variance-covariance \( DS(\theta_0)^{-1/2} \hat{S}_n(\theta)^{-1/2} G_n^* (\theta) \), with \( DS(\theta_0) \) being the diagonal submatrix of \( S(\theta_0) \).

**Comment.** The key difference between the test statistics \( Q^c_n(\theta) \) and \( \operatorname{MaxS}_n(\theta) \) is how the bias-corrected moment function \( G_n^* (\theta) \) is normalized: the former uses the entire variance-covariance matrix, while the latter only uses the diagonal elements of the variance-covariance matrix for scaling. The statistic \( Q^c_n(\theta) \) is considered by Stock and Wright (2000). The statistic \( \operatorname{MaxS}_n(\theta) \) is

---

6 Andrews and Soares (2010) consider inference for a general moment equality/inequality model. We specialize their method to a case with only moment equalities, where “moments” should be interpreted as integrated stochastic quantities (see (2.4)).
a special case of Andrews and Soares (2010); we use this special case only for concreteness, noting that the result can be easily extended to other test statistics considered there.

Let \( \alpha \in (0, 1) \). A nominal level \( 1 - \alpha \) confidence set can be constructed by collecting all \( \theta \in \Theta \) for which the null hypothesis \( H_0 : \theta_0 = \theta \) is not rejected at significance level \( \alpha \). Let \( \chi^2_{k,1-\alpha} \) be the \( 1 - \alpha \) quantile of a \( \chi^2_k \) distribution. Proposition \( 2(a) \) implies that the set \( CS_{n,1-\alpha}^{SW} \equiv \{ \theta \in \Theta : \Delta_n^{-1}Q_n^c(\theta) \leq \chi^2_{k,1-\alpha} \} \) has valid asymptotic coverage, i.e., \( P(\theta_0 \in CS_{n,1-\alpha}^{SW}) \rightarrow 1 - \alpha \). The asymptotic distribution of MaxS\(_n\) (\( \theta \)) is not pivotal. Its critical value can be obtained as follows. First, estimate the asymptotic variance-covariance matrix of \( \tilde{\xi} \) in Proposition \( 2(b) \) by \( \hat{\Sigma}_{\xi,n} = \hat{D}S_n(\theta)^{-1/2} \hat{S}_n(\theta) \hat{D}S_n(\theta)^{-1/2} \). Second, draw a large number of Monte Carlo realizations of \( \tilde{\xi} \) from \( N(0, \hat{\Sigma}_{\xi,n}) \). Third, compute the Monte Carlo \( 1 - \alpha \) quantile of \( M(\tilde{\xi}) \), denoted by \( cv_{n,1-\alpha}^{AS}(\theta) \).

By Theorem \( 3 \) and the continuous mapping theorem, \( cv_{n,1-\alpha}^{AS}(\theta_0) \) consistently estimates the \( 1 - \alpha \) \( F \)-conditional quantile of the limit distribution of MaxS\(_n\) (\( \theta_0 \)). A \( 1 - \alpha \) nominal level confidence set can be constructed as

\[
CS_{n,1-\alpha}^{AS} = \{ \theta \in \Theta : MaxS_n(\theta) \leq cv_{n,1-\alpha}^{AS}(\theta) \},
\]

which has valid asymptotic coverage by Proposition \( 2(b) \), i.e., \( P(\theta_0 \in CS_{n,1-\alpha}^{AS}) \rightarrow 1 - \alpha \).

### 2.7 An alternative bias-correction procedure

We now discuss an alternative bias-correction procedure. The idea is to directly correct the bias in the raw estimator \( \hat{\theta}_n \). In view of Proposition \( 1 \), we consider the following estimator

\[
\theta_n^* \equiv \hat{\theta}_n + \frac{1}{k_n}(\hat{H}_n(\hat{\theta}_n)\hat{\Sigma}_n\hat{H}_n(\hat{\theta}_n))^{-1}\hat{H}_n(\hat{\theta}_n)\hat{\Sigma}_n\hat{B}_n(\hat{\theta}_n),
\]  \(\text{(2.8)}\)

which we refer to as the two-step estimator.

**Theorem 5** We have \( \Delta_n^{-1/2}(\theta^*_n - \theta_0) = M\xi_n + o_p(1) \), where \( M \) and \( \xi_n \) are the same as in Theorem 2.

**Comments.** (i) Theorem 5 shows that the two-step estimator \( \theta^*_n \) has the same asymptotic representation as the one-step estimator \( \theta^*_n \). In this sense, these estimators are asymptotically equivalent.

(ii) The preliminary estimator in the correction term in \( (2.8) \) does not need to be \( \hat{\theta}_n \). In particular, one can iterate the bias-correction by setting \( \theta_n^{(1)} = \theta^*_n \) and for \( j \geq 2 \), \( \theta_n^{(j)} = \hat{\theta}_n + k_n^{-1}(\hat{H}_n(\theta_n^{(j-1)})\hat{\Sigma}_n\hat{H}_n(\theta_n^{(j-1)}))^{-1}\hat{H}_n(\theta_n^{(j-1)})\hat{\Sigma}_n\hat{B}_n(\theta_n^{(j-1)}) \). With the same proof, it can be shown that \( \Delta_n^{-1/2}(\theta_n^{(j)} - \theta_0) = M\xi_n + o_p(1) \) for \( j \geq 2 \).
Conducting the bias-correction within the estimation step is clearly a desirable feature for the one-step estimator $\theta_n^\star$ and the continuous updating estimator $\theta_c^n$. Moreover, as shown in Section 3 below, the bias-corrected moment function $G_n^\star$ and objective functions $Q_n^\star$ and $Q_c^n$ can be used for overidentification test. That said, we consider the two-step estimator $\theta_n^\ast$ an interesting complement to its one-step counterpart. An in-depth comparison between these estimators via higher-order asymptotic expansion is left for future research.

3 Testing

3.1 Overidentification test

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

**Theorem 6** Suppose $\Xi = S(\theta_0)^{-1}$. Let $\theta_n^\ast$ and $\theta_n^\star$ be given as in Theorems 2 and 5. Then

$$\Delta_n^{-1}Q_n^\star(\theta_n^\ast) \xrightarrow{d} \chi^2_{k-\dim(\theta)}$$

and the same convergence holds for $\Delta_n^{-1}Q_n^\star(\theta_n^\star)$ and $\Delta_n^{-1}Q_c^n(\theta_c^n)$.

**Comment.** The $J$ statistic is evaluated with the bias-corrected objective function $Q_n^\star$. The “raw” objective function $\hat{Q}_n$ cannot be used for such purpose. Indeed, it can be shown that $k_n^2\hat{Q}_n(\tilde{\theta}_n) \xrightarrow{p} B^\top \Xi B$ for any $\tilde{\theta}_n$ satisfying $\tilde{\theta}_n \xrightarrow{p} \theta_0$.

3.2 Parameter stability

We now consider a multi-period problem. The time span of period $\tau$, $\tau = 1, \ldots, \bar{\tau}$, is $[(\tau - 1)T, \tau T]$ with $T > 0$ and $\bar{\tau} \geq 2$ fixed.\(^7\) We assume that the true parameter $\theta_0(\tau)$ is constant for each $\tau$ and may vary across periods. We are interested in constructing the uniform confidence band for the parameter path $\tau \mapsto \kappa^\top \theta_0(\tau)$, where $\kappa = (\kappa_j)$ is a $\dim(\theta) \times 1$ constant vector. We denote by $\theta_n^\ast(\tau)$ the one-step estimator based on period-$\tau$ data and denote by $\hat{\Sigma}_n(\tau)$ a consistent estimator for $\Sigma(\tau)$, with $\Sigma(\tau)$ being the asymptotic variance associated with $\theta_n^\ast(\tau)$; see Theorems 2 and 3. For $\alpha \in (0, 1)$, let $z_{\tau,\alpha}$ denote the $1 - \alpha/2$ quantile of the variable $\max_{1 \leq \tau \leq \bar{\tau}} |N_\tau|$, where $N_\tau$, $1 \leq \tau \leq \bar{\tau}$, are independent standard normal variables. A nominal level $1 - \alpha$ uniform confidence band for $\kappa^\top \theta_0(\tau)$ is given by the following sequence of interval estimators: for $1 \leq \tau \leq \bar{\tau}$,

$$CI_n(\tau; \kappa, 1 - \alpha) \equiv \left[ \kappa^\top \theta_n^\ast(\tau) - z_{\tau,\alpha} \Delta_n^{1/2} \sqrt{\kappa^\top \hat{\Sigma}_n(\tau) \kappa}, \kappa^\top \theta_n^\ast(\tau) + z_{\tau,\alpha} \Delta_n^{1/2} \sqrt{\kappa^\top \hat{\Sigma}_n(\tau) \kappa} \right].$$

It is easy to see that similar constructions can be based on the continuous updating estimator and the two-step estimator. We omit the details for brevity. The asymptotic property of the above

\(^7\)The assumption that the time horizon is constant across periods is not essential and can be easily relaxed.
confidence band follows the asymptotic independence of \( \theta_n^*(\tau) \) across periods, as is typical in this type of problems involving high-frequency data.

**Corollary 1** The sequence of variables \( \Sigma(\tau)^{-1/2}(\theta_n^*(\tau) - \theta_0(\tau)) \), \( 1 \leq \tau \leq \bar{\tau} \), converges in distribution to a \( \bar{\tau} \)-dimensional standard normal distribution. Moreover, for any \( \alpha \in (0, 1) \) and \( \dim(\theta) \times 1 \) vector \( \kappa \neq 0 \),

\[
P(\kappa^T \theta_0(\tau) \in CI_n(\tau; \kappa, 1 - \alpha)) \text{ for all } 1 \leq \tau \leq \bar{\tau} \rightarrow 1 - \alpha.
\]

The uniform confidence band is a natural tool for visualizing the temporal stability or instability of the one-dimensional parameter \( \kappa^T \theta_0(\tau) \). In practice, one may also be interested in testing for the joint stability of the vector \( \theta_0(\tau) \). Formally, we consider the following hypotheses:

\[
\begin{align*}
H_0 &: \theta_0(1) = \cdots = \theta_0(\bar{\tau}) \\
H_a &: \theta_0(\tau) \neq \theta_0(\tau') \text{, some } 1 \leq \tau, \tau' \leq \bar{\tau}.
\end{align*}
\]

(3.1)

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

\[
R = \begin{pmatrix}
1 & -1 \\
-1 & 1 \\
\ddots & \ddots \\
1 & -1
\end{pmatrix}^{(\bar{\tau} - 1) \times \bar{\tau}} \otimes I_{\dim(\theta)}
\]

where \( \otimes \) is the Kronecker product and \( I_{\dim(\theta)} \) is the \( \dim(\theta) \)-dimensional identity matrix, and

\[
\Sigma_n = \begin{pmatrix}
\widehat{\Sigma}_n(1) \\
\ddots \\
\widehat{\Sigma}_n(\bar{\tau})
\end{pmatrix}, \quad \Sigma = \begin{pmatrix}
\Sigma(1) \\
\ddots \\
\Sigma(\bar{\tau})
\end{pmatrix}.
\]

Notice that \( (\Delta \theta_n^*(1), \ldots, \Delta \theta_n^*(\bar{\tau} - 1)) \xrightarrow{\mathcal{L}} \mathcal{M}(0, R\Sigma R^T) \) by Corollary 1. Let \( \mathbb{M}_+ \) be the space of positive definite \( \dim(\theta)\bar{\tau} \times \dim(\theta)\bar{\tau} \) matrices. We consider test statistics of the following form:

\[
PS_n = S(\Delta \theta_n^*(1), \ldots, \Delta \theta_n^*(\bar{\tau} - 1), \Sigma_n),
\]

where the test function \( S : \mathbb{R}^{(\bar{\tau} - 1)\dim(\theta)} \times \mathbb{M}_+ \mapsto \mathbb{R}_+ \) satisfies the following properties: (i) \( S(\cdot) \) is continuous; (ii) \( S(x_1, \ldots, x_{\bar{\tau} - 1}; \Sigma) = 0 \) if and only if \( x_\tau = 0 \) for all \( 1 \leq \tau \leq \bar{\tau} - 1 \); (iii) \( S(x_1, \ldots, x_{\bar{\tau} - 1}; \Sigma) \rightarrow \infty \) whenever \( \|x_\tau\| \rightarrow \infty \) for some \( 1 \leq \tau \leq \bar{\tau} - 1 \). We list two examples for
concreteness: 
\[
S(\Delta \theta_n^*(1), \ldots, \Delta \theta_n^*(\bar{\tau} - 1); \tilde{\Sigma}_n) = \left\{ \begin{array}{ll}
\max_{1 \leq \tau \leq \bar{\tau} - 1} \| (\hat{D}_n(\tau) + \hat{D}_n(\tau + 1))^{-1/2} \Delta \theta_n^*(\tau) \| \\
\max_{1 \leq \tau \leq \bar{\tau} - 1} \Delta \theta_n^*(\tau)^T (\tilde{\Sigma}_n(\tau) + \tilde{\Sigma}_n(\tau + 1))^{-1} \Delta \theta_n^*(\tau),
\end{array} \right.
\]
where \( \hat{D}_n(\tau) \) collects the diagonal elements of \( \tilde{\Sigma}_n(\tau) \).

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

**Corollary 2** Consider the hypotheses in (3.1). We have the following.

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

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

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

Corollary 2(a) describes the asymptotic distribution of \( PS_n \) under the null hypothesis. A nominal level \( \alpha \) test for parameter stability can be carried out by comparing \( PS_n \) with a consistent estimator of the \( 1 - \alpha \) \( F \)-conditional quantile of the limit distribution; we reject \( H_0 \) for large \( PS_n \). Since the limit null distribution is nonstandard, there is no analytical expression for the quantile in general. Corollary 2(b) shows that the \( 1 - \alpha \) \( F \)-conditional quantile of \( S((R\tilde{\Sigma}_n R^\top)^{1/2} \zeta; \tilde{\Sigma}_n) \) is a consistent estimator for the quantile of the limit distribution; this estimator can be computed via simulation by drawing a large number of Monte Carlo realizations of \( \zeta \) with \( \tilde{\Sigma}_n \) given. Since the critical values form a tight sequence, part (c) of the corollary implies that the aforementioned test has asymptotic power one under \( H_a \).

4 Simulation results

In this section, we examine the asymptotic theory above in two simulation settings that mimic the setup of our empirical application in Section 5. Throughout the simulations, we fix \( T = 63 \) days and consider two sampling frequencies: \( \Delta = 15 \) or 60 seconds. 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.
4.1 Simulation I: affine jump-diffusion volatility dynamics

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

\[
\begin{align*}
    dX_t &= (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t}dW_t + J_X dN_t \\
    dV_t &= \kappa(\bar{v} - V_t)dt + \sigma \sqrt{V_t}dB_t + J_V dN_t \\
\end{align*}
\]

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 $J_X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $J_V \sim \exp(\beta_V)$. We set $\kappa = 5$, $\sigma = 0.75$, $\rho = -0.8$, $\bar{v} = 0.06$, $\mu_X = -0.02$, $\sigma_X = 0.05$, $\beta_V = 0.05$, $\lambda_0 = 30$, $\lambda_1 = 60$, $\mu_0 = 0.05$, and $\mu_1 = 0.5$. We simulate $Y_t$ according to

\[
Y_t = a + b V_t + \varepsilon_t,
\]

where $\varepsilon_t \sim \text{IID} \mathcal{N}(0, \sigma_\varepsilon^2)$, $a = 0.056$, $b = 0.631$, and $\sigma_\varepsilon = 0.03$. This model is studied in Section 5 for pricing the VIX; the parameter values are calibrated to our data.

Figure 1 presents finite-sample distributions of the raw and the one-step estimators. In line with the asymptotic theory, the raw estimators exhibit large (relative to their sampling variability) biases, while the one-step estimators are properly centered at the true values. This finding is further confirmed by Table 1, which compares the biases and the root-mean-squared errors of the raw and the one-step estimates.

Figure 2 compares finite-sample distributions of the studentized raw and one-step estimates. Although the raw estimator does not admit a CLT, we studentize it with the “naive” standard error, which is computed according to the standard linear least-squares theory while treating $V_t$ as if it were estimated without error. As predicted by the asymptotic theory, the distribution of the studentized one-step estimator is well approximated by the $\mathcal{N}(0, 1)$ distribution. On the contrary, the distribution of the studentized raw estimator differs substantially from $\mathcal{N}(0, 1)$.

4.2 Simulation II: exponential-OU volatility dynamics

Following Chernov, Gallant, Ghysels, and Tauchen (2003) (see also Huang and Tauchen (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Gonçalves and Meddahi (2009) for similar models without jumps), we simulate $X_t$ and $V_t$ according to

\[
\begin{align*}
    dX_t &= (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t}dW_t + J_X dN_t - \mu_X \lambda dt, \\
    \log V_t &= \alpha + \beta F_t, \\
    dF_t &= -\kappa F_t dt + \sigma dB_t + J_F dN_t - \mu_F \lambda dt, \\
\end{align*}
\]

Following Chernov, Gallant, Ghysels, and Tauchen (2003) (see also Huang and Tauchen (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Gonçalves and Meddahi (2009) for similar models without jumps), we simulate $X_t$ and $V_t$ according to

\[
\begin{align*}
    dX_t &= (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t}dW_t + J_X dN_t - \mu_X \lambda dt, \\
    \log V_t &= \alpha + \beta F_t, \\
    dF_t &= -\kappa F_t dt + \sigma dB_t + J_F dN_t - \mu_F \lambda dt, \\
\end{align*}
\]

(4.3)
Table 1: Summary of results in Simulation I

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>Root-Mean-Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td>One-Step</td>
</tr>
<tr>
<td></td>
<td>$\Delta = \bar{\kappa} = 0$</td>
<td>$\bar{\kappa} = 1$</td>
</tr>
<tr>
<td></td>
<td>$\bar{\kappa} = 0.5$</td>
<td>0.009</td>
</tr>
<tr>
<td>a 15 sec</td>
<td></td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b 15 sec</td>
<td>-0.049</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>-0.083</td>
<td>-0.066</td>
</tr>
<tr>
<td>60 sec</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: We report the bias and the root-mean-squared error of the raw and the one-step estimators in Simulation I, where the regression model is $Y_t = a + bV_t + \varepsilon_t$. We set $T = 63$ days, the sampling interval $\Delta = 15$ or 60 seconds, and $k_n = [\bar{\kappa}^{2/5} \Delta^{-2/5}]$. The true parameter values are $a = 0.056$ and $b = 0.631$.

with $\mathbb{E}[dW_t dB_t] = \rho dt$, $J_X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $J_F \sim \mathcal{N}(\mu_F, \sigma_F^2)$, and $N_t$ being a Poisson process with intensity $\lambda$. We set $\alpha = -2.80$, $\beta = 6.25$, $\rho = -0.75$, $\mu_X = -0.02$, $\sigma_X = 0.05$, $\mu_F = 0.02$, $\sigma_F = 0.02$, $\sigma = 1$, $\kappa = 4$, $\lambda = 25$, $\mu_0 = 0.05$ and $\mu_1 = 0.50$. This data generating process differs from (4.1) in an important way in that the spot variance process follows an exponential-OU model.

We simulate $Y_t$ according to

$$Y_t = a + bV_t^c + \varepsilon_t,$$

where $\varepsilon_t \sim \mathcal{IH}(0, \sigma_\varepsilon^2)$ and the parameters $a = 0.0173$, $b = 0.4732$, $c = 0.5796$ and $\sigma_\varepsilon = 0.02$ are calibrated to the dataset employed in Section 5. We refer to (4.4) as the “ABC” model. This model is a parsimonious approximation to a fairly complicated nonlinear VIX pricing model implied by risk-neutral dynamics similar to (4.3); see Section 5 for details. The parameter of interest is now $\theta = (a, b, c)$.

We estimate the model (4.4) via nonlinear least-squares by setting $d_j(\cdot) = \partial_\theta f_j(\cdot)$ and conduct bias-correction using the one-step estimator. The results are organized similarly as in Section 4.1.
Figure 1: Histograms of nonstudentized estimators in Simulation I

Note: This figure compares the finite-sample distributions of the raw estimators (solid) and the one-step estimators (shaded area) in Simulation I, where the regression model is $Y_t = a + bV_t + \epsilon_t$. The dashed lines highlight the true parameter values. The sampling interval is $\Delta = 15$ seconds (left) or 60 seconds (right). We set $T = 63$ days and $k_n = [\Delta^{-2/5}]$. There are 2,000 Monte Carlo trials.

We summarize our findings as follows. Figure 3 and Table 2 show that the bias-correction is quite effective. Figure 4 shows that the finite-sample distributions of the studentized one-step estimators are much closer to $\mathcal{N}(0, 1)$ than their “naive” counterparts; the latter are computed based on the standard nonlinear least-squares theory. However, we do observe that the distributions of the one-step estimators deviate from the asymptotic normal approximation to a nontrivial extent. Our conjecture is that the exponent parameter in the ABC model is weakly identified, leading to finite-sample distortions to all three parameters (Stock and Wright (2000)). In Section 4.3 below, we further investigate this issue in a testing context.

4.3 Tests for the linear specification

In our empirical application, we nest the linear model (4.2) in the ABC model (4.4) and conduct specification tests for the null hypothesis $H_0 : c = 1$. Such a test can be carried out by first estimating the ABC model and then conducting the t-test. Since the t-test relies on the asymptotic normal approximation, one may expect the test to suffer from size distortions in view of Figure 4. Alternatively, we can test the null hypothesis by examining whether the Anderson-Rubin-type
Table 2: Summary of Results in Simulation II

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>Raw</th>
<th>One-Step</th>
<th>Raw</th>
<th>One-Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta)</td>
<td>(\hat{\kappa} = 0.5)</td>
<td>(\hat{\kappa} = 1)</td>
<td>(\hat{\kappa} = 2)</td>
<td>(\hat{\kappa} = 0.5)</td>
<td>(\hat{\kappa} = 1)</td>
</tr>
<tr>
<td>(a)</td>
<td>15 sec</td>
<td>-0.007</td>
<td>-0.006</td>
<td>-0.005</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>-0.013</td>
<td>-0.011</td>
<td>-0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>(b)</td>
<td>15 sec</td>
<td>-0.022</td>
<td>-0.018</td>
<td>-0.016</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>-0.033</td>
<td>-0.028</td>
<td>-0.025</td>
<td>0.000</td>
</tr>
<tr>
<td>(c)</td>
<td>15 sec</td>
<td>-0.045</td>
<td>-0.037</td>
<td>-0.031</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>-0.072</td>
<td>-0.061</td>
<td>-0.053</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Root-Mean-Squared-Error</th>
<th>Raw</th>
<th>One-Step</th>
<th>Raw</th>
<th>One-Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta)</td>
<td>(\hat{\kappa} = 0.5)</td>
<td>(\hat{\kappa} = 1)</td>
<td>(\hat{\kappa} = 2)</td>
<td>(\hat{\kappa} = 0.5)</td>
<td>(\hat{\kappa} = 1)</td>
</tr>
<tr>
<td>(a)</td>
<td>15 sec</td>
<td>0.012</td>
<td>0.010</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>0.022</td>
<td>0.019</td>
<td>0.017</td>
<td>0.010</td>
</tr>
<tr>
<td>(b)</td>
<td>15 sec</td>
<td>0.036</td>
<td>0.031</td>
<td>0.027</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>0.056</td>
<td>0.050</td>
<td>0.045</td>
<td>0.030</td>
</tr>
<tr>
<td>(c)</td>
<td>15 sec</td>
<td>0.055</td>
<td>0.047</td>
<td>0.041</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>60 sec</td>
<td>0.090</td>
<td>0.079</td>
<td>0.073</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Note: We report the bias and the root-mean-squared error of the raw and the one-step estimators in Simulation II, where the regression model is \(Y_t = a + bV_t^c + \varepsilon_t\). We set \(T = 63\) days, the sampling interval \(\Delta = 15\) or 60 seconds, and \(k_n = [\hat{\kappa}^{2/5}\Delta^{-2/5}]\). The true parameter values are \(a = 0.0173\), \(b = 0.4732\) and \(c = 0.5796\).
Figure 2: Histograms of studentized estimators in Simulation I

Intercept, 15s

Intercept, 60s

Slope, 15s

Slope, 60s

Note: This figure compares the finite-sample distributions of the studentized raw estimators (solid) and one-step estimators (shaded area) in Simulation I, where the regression model is $Y_t = a + bV_t + \varepsilon_t$. Plots for the raw estimators are hardly discernible as they are beyond the limits of the axes. The $N(0, 1)$ density function is plotted for comparison (dashed). The sampling interval is $\Delta = 15$ seconds (left) or 60 seconds (right). We set $T = 63$ days and $k_n = [\Delta^{-2/5}]$. There are 2,000 Monte Carlo trials.

Table 3 reports finite-sample rejection rates of the t-tests using the raw and the one-step estimators, as well as coverage rates of $CS_{n,1-\alpha}^{SW}$ and $CS_{n,1-\alpha}^{AS}$ for the true parameter vector. Note that the coverage rates of the Anderson-Rubin-type confidence sets provide upper bounds for rejection rates of the associated tests for $H_0 : c = 1$. The findings are as follows. First, the raw test, i.e. the t-test using the raw estimator, almost always falsely rejects the null hypothesis. Second, the one-step test rejects much less than the raw test, but still suffers from considerable over-rejection. The over-rejection is mitigated at higher sampling frequency, suggesting that this is a finite-sample distortion. Third, coverage rates of $CS_{n,1-\alpha}^{SW}$ and $CS_{n,1-\alpha}^{AS}$ are close to the nominal level. In particular, the coverage rate of $CS_{n,1-\alpha}^{AS}$ is either below or very close to the nominal level, suggesting that the associated test has good size control, although it may be conservative. We thus
Figure 3: Histograms of nonstudentized estimators in Simulation II

Note: This figure compares the finite-sample distributions of the raw estimators (solid) and the one-step estimators (shaded area) in Simulation II, where the regression model is $Y_t = a + bV_t^c + \varepsilon_t$. The dashed lines highlight the true parameter values. The sampling interval is $\Delta = 15$ seconds (left) or 60 seconds (right). We set $T = 63$ days and $k_n = [\Delta^{-2/3}]$. There are 2,000 Monte Carlo trials.

consider the AS test as a conservative alternative to the t-test based on the one-step estimator.

5 An empirical application on VIX pricing models

5.1 Setup

Model specification is one of the central topics in the empirical option pricing literature. We apply the econometric method above to investigate the specification of the risk-neutral dynamics of the stochastic volatility process by using intraday data of the S&P 500 index and the VIX.\(^8\) The sample

\(^8\)The VIX is constructed from a portfolio of S&P 500 options. In theory, the squared VIX, up to a scaling constant, is approximately equal to the risk-neutral expectation of the quadratic variation of log returns of the S&P
Figure 4: Histograms of studentized estimators in Simulation II

Note: This figure compares the finite-sample distributions of the studentized raw estimators (solid) and one-step estimators (shaded area) in Simulation II, where the regression model is $Y_t = a + bV_t^c + \varepsilon_t$. The $\mathcal{N}(0, 1)$ density function is plotted for comparison (dashed). The sampling interval is $\Delta = 15$ seconds (left) or 60 seconds (right). We set $T = 63$ days and $k_n = [\Delta^{-2/5}]$. There are 2,000 Monte Carlo trials.
Table 3: Comparison of Finite-Sample Null Rejection Rates

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>One-Step</th>
<th>SW</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>6.95</td>
<td>2.45</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>15.90</td>
<td>9.00</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>24.45</td>
<td>16.05</td>
<td>9.60</td>
</tr>
</tbody>
</table>

$\Delta = 15$ seconds

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>One-Step</th>
<th>SW</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>4.40</td>
<td>1.85</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>12.70</td>
<td>8.00</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>19.95</td>
<td>14.00</td>
<td>8.70</td>
</tr>
</tbody>
</table>

Note: We report finite-sample null rejection rates (%) of 1%, 5% and 10% nominal level tests for $H_0: c = 1$. We consider two-sided t-tests based on the raw estimator (Raw) and the one-step bias-corrected estimator (One-Step). We also report coverage rates of the Andersen-Rubin-type confidence sets $CS_{n,1-\alpha}^{SW}$ (SW) and $CS_{n,1-\alpha}^{AS}$ (AS) for $\theta_0$, which provide upper bounds for the rejection rates of the associated tests. We set $T = 63$ days, the sampling interval is $\Delta = 15$ or 60 seconds, and $k_n = \lceil \Delta^{-2/5} \rceil$.

period ranges from January 2007 to September 2012, as constrained by data availability; the data source is TickData Inc.. The VIX is announced by the CBOE at roughly every 15 seconds. The S&P 500 index data is sampled more frequently. In order to reduce the effect of microstructure noise and asynchronicity, we resample the data at every minute. Days with trading hours shorter than regular trading hours are eliminated, resulting in a sample of 1,457 days spanning 23 quarters. We split two quarters (2007 Q3 and 2011 Q3) into halves as their subsample estimates tend to be more stable; see Section 5.2 for details. Our final sample consists of 25 subsample periods, including 21 quarters and 4 half-quarters and each period is treated on its own. The diurnal pattern of the intraday spot variance movement is adjusted according to standard procedures in the literature (see e.g. Todorov and Tauchen (2012)).

The dynamics of the logarithm of the index $X_t$ under the risk-neutral measure, i.e. the “$Q$ measure”, is assumed to follow

$$X_t = X_0 + \int_0^t b_s^Q \, ds + \int_0^t \sqrt{V_s^Q} \, dW_s^Q + \int_0^t \int_R z (N(ds,dz) - \nu(V_s,dz)) \, ds$$

where the drift $b_s^Q$ is determined by the no-arbitrage condition and is irrelevant for the pricing of VIX, $W^Q$ is a $Q$ Brownian motion, and $N(ds,dz)$ is the jump measure of $X_t$ with compensator 500 index (Jiang and Tian (2005), Carr and Wu (2009)).
\(\nu(V_s, dz)ds\) that may depend on the spot variance. We assume there exist constants \(\eta_0\) and \(\eta_1\), such that for all \(v > 0\),
\[
\int_{\mathbb{R}} z^2 \nu(v, dz) = \eta_0 + \eta_1 v.
\]
This assumption is satisfied by most models in the empirical option pricing literature (Singleton (2006)) and is maintained throughout our empirical application.

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

**Type-I Model:**
\[
V_t = V_0 + \int_{0}^{t} \kappa(\bar{v} - V_s)ds + M^Q_t
\]  
where \(M^Q\) is a martingale under the \(Q\)-measure and \(\kappa, \bar{v}\) are model parameters. Type-I models include those studied by Bakshi, Cao, and Chen (1997), Bates (2000), Pan (2002), Eraker (2004), Eraker, Johannes, and Polson (2003), Broadie, Chernov, and Johannes (2007), Song and Xiu (2012), and Bates (2012), among others, where jumps may be driven by the compound Poisson process with time-varying intensity or the CGMY process (Carr, Geman, Madan, and Yor (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.

Below, we denote \(Y_t = \left(\text{VIX}_t / 100\right)^2\), which is referred to as the implied variance, and denote by \(Y^*_t\) the associated theoretical value, in compliance with the notational convention (2.1).

**Proposition 3** Type-I models imply the following pricing formula for the implied variance
\[
Y^*_t = a + bV_t,
\]  
where \(a\) and \(b\) only depend on model parameters (see the proof for details).

Proposition 3 highlights the advantage for using the VIX to study the risk-neutral volatility dynamics. The large class of Type-I models, although potentially very different from each other with distinct pricing implications for individual options, all imply the linear pricing function for the implied variance. Consequently, specification tests for this large class of structural models can be carried out by examining the specification of the linear regression model (4.2). We do so by nesting (4.2) as a special case of the ABC model (4.4) and testing whether \(c = 1\).

We adopt the ABC model as the encompassing alternative for two reasons. First, it also
“almost” nests the log-linear model as a special case with $a = 0$. Indeed, the ABC model is the most parsimonious model encompassing both linear and log-linear specifications and hence, at least in our opinion, should be of great practical interest.

Second, the ABC model can be considered as an approximation for a more complicated non-linear structural model. To illustrate this point, we consider Type-II models which impose an exponential-affine structure for the $Q$-dynamics of $V_t$:

$$
\text{Type-II Model: } \log V_t = \alpha + \beta F_t, \quad F_t = F_0 - \int_0^t \kappa F_s ds + L_t^Q,
$$

where $L_t^Q$ is a finite variational Lévy martingale with diffusive coefficient $\sigma$ and Lévy measure $\tilde{\nu}$, and $\alpha, \beta, \kappa$ are model parameters. This model dates back to Nelson (1990), who introduces it as a continuous-time limit of the discrete EGARCH model. Andersen, Bollerslev, and Meddahi (2005) and Chernov, Gallant, Ghysels, and Tauchen (2003) have employed this model in their empirical work.

**Proposition 4** (a) Type-II models imply the following pricing formula for the implied variance

$$
Y_t^* = \eta_0 + \frac{1}{\tau} \int_0^\tau (\eta_1 + 1) \exp \left( \alpha + e^{-\kappa u}(\log V_t - \alpha) + C(u) \right) du,
$$

where $\tau = 21$ trading days is the horizon of the VIX, $C(u) = \int_0^u \varphi(\beta e^{-\kappa v}) dv$ and $\varphi(u) = \sigma^2 u^2/2 + \int_\mathbb{R} (e^{uz} - 1 - uz)\tilde{\nu}(dz)$, provided that $\varphi(u) < \infty$ for all $u$ bounded between $\beta$ and $\beta e^{-\kappa\tau}$, and that $\mathbb{E} [\log(1 + \max\{-L_{t\tau}^Q, 0\})] < \infty$. (b) $Y_t^* = a + bV_t + O(\tau)$ for $a = \eta_0$ and $b = \eta_1 + 1$, as $\tau \to 0$. (c) $Y_t^* = a(\tau) + b(\tau) V_t^{c(\tau)} + O(\tau^2)$ for $a(\tau) = \eta_0$, $b(\tau) = (\eta_1 + 1)e^{\alpha - \alpha e^{-\kappa\tau/2} + C(\tau/2)}$ and $c(\tau) = e^{-\kappa\tau/2}$, as $\tau \to 0$.

**Comments.** (i) Proposition 4(a) shows the exact functional relationship between the theoretical price $Y_t^*$ and the spot variance $V_t$. This pricing formula is similar to that in Todorov and Tauchen (2011).

(ii) Part (b) and part (c) show that, when the VIX horizon $\tau$ is considered “small”, the linear specification is the leading approximation term of the exact pricing formula (5.3), and the ABC model provides a higher-order approximation. Similar expansions have been considered by Medvedev and Scaillet (2007) and Xiu (2012) for option implied volatilities and prices, respectively.

(iii) To the extent that the ABC model is a good approximation to the pricing formula (5.3), its exponent parameter $c$ has a structural interpretation: $c = \exp(-\kappa\tau/2)$. In particular, $c < 1$
(resp. $c > 1$) is equivalent to $\kappa > 0$ (resp. $\kappa < 0$), i.e. the volatility process is mean-reverting (resp. mean-repelling).

We caution the reader that the ABC model is only an approximation, rather than a substitute, to the nonlinear pricing function (5.3). We argue that this simple reduced-form model is close to an important class of structural models in a well-defined sense of approximation, and hence serves as an economically interesting and statistically parsimonious alternative for testing against the linear pricing model (4.2).

5.2 Results

For each of the 25 subsample periods, we estimate the ABC model via nonlinear least-squares using the one-step estimator. In view of comment (iii) of Proposition 4, we preclude mean-repelling volatility dynamics by restricting $c \leq 1$ in our estimation. Figure 5 shows the scatter plots of the implied variance versus the spot variance estimate, along with the fitted ABC pricing functions, for two representative quarters. For comparison, we also plot the fitted pricing function using the raw nonlinear least-squares estimator. In both quarters, we observe evident concavity in the pricing function, suggesting that the linear regression model is likely to be misspecified. Figure 6 presents similar plots for the two split quarters, i.e. 2007 Q3 and 2011 Q3. Notice that the two subsamples in each quarter exhibit quite distinct dependence between the implied variance and the spot variance, which is our initial motivation for sample splitting.

Figure 7 plots the one-step estimates for the linear regression model and the ABC model across subsample periods, along with the 95% uniform confidence bands. Our findings are summarized as follows. First, parameter estimates in the linear regression model exhibit statistically significant time variation, especially for the intercept. Hence, any Type-I model with constant parameters over the entire sample period will fail to capture the dependence of the implied variance on the spot variance. Second, parameter paths of the ABC model also exhibit significant time variation, but to a much smaller extent than those in the linear regression model. Third, among 25 subsample periods, the exponent parameter $c$ in the ABC model is significantly less than 1 for 20 (resp. 18) periods at the 5% (resp. 1%) nominal level, suggesting that the linear model is likely to be misspecified. This finding of course needs to be interpreted with caution in view of the size distortion shown in Table 3, but it is unlikely that this finding is solely due to size distortion. We further conduct the AS test for the null hypothesis $H_0 : c = 1$. At the 5% significance level, the AS test rejects the null for only one quarter, 2008 Q4—the peak of the recent financial crisis. The null hypothesis is also rejected in this quarter at the 1% nominal level. The lack of rejection of the AS test is not surprising as this type of test is known to be conservative with low power (Andrews and Soares (2010), Bugni, Canay, and Shi (2013)). The rejection of $H_0 : c = 1$ by this conservative test during the crisis period should be interpreted as strong evidence against the linear
Figure 5: Fitted ABC Models

Note: This figure shows scatter plots for 2 representative quarters 2007 Q4 and 2012 Q2, along with fitted ABC pricing functions using the one-step estimator (solid) and the raw estimator (dot-dashed). The sampling interval is $\Delta = 1$ minute, and $k_n = [\Delta^{-2/5}]$.

pricing model. Clearly, the substantial variation in the volatility level during the crisis period helps improve the discriminant power of the specification test.

Overall, we find strong statistical evidence for parameter instability in the linear regression model. Parameter instability is mitigated, but not eliminated, by the more general ABC model. Specification tests within each period suggest that the linear VIX pricing model is likely to be misspecified over the quarterly horizon.

6 Concluding remarks and future works

We study the inference of a general continuous-time nonlinear vector regression model with the latent spot variance as a covariate. We demonstrate the higher-order bias, and hence the lack of CLT, of the raw estimator. We also propose bias-correction methods and inference tools. We illustrate our method via a novel empirical application using high-frequency VIX and S&P 500 index data. We examine the implications of a general class of risk-neutral volatility models (Type-I models) for VIX pricing, and provide statistical evidence against these conventional models. Constructively, we propose the ABC model for VIX pricing, which may be used as a benchmark in future empirical work.

In view of the numerous theoretical complications and refinements in GMM, or moment equality models in general, we restrict the goal of this paper mainly to establishing the basic inference results for high-frequency spot variance regressions. Much remains to be done. In par-
Figure 6: Fitted ABC Models for Split Quarters

Note: This figure shows scatter plots for the two split quarters, 2007 Q3 and 2011 Q3, along with fitted ABC pricing functions using the one-step estimator (solid) and the raw estimator (dot-dashed). Light-gray dots (resp. dark-gray circles) are data points in the first (resp. second) half quarter. The sampling interval is $\Delta = 1$ minute, and $k_n = [\Delta^{-2/5}]$.

In particular, the finite-sample distortions in the estimation of nonlinear models call for further improvement. The weak-identification perspective (Stock and Wright (2000), Andrews and Cheng (2012)) should be helpful in this regard. An in-depth study in this direction seems to require its own paper. Other extensions include nonparametric regressions (Chen (2007), Härdle and Linton (1994)), extreme estimation for nonsmooth objective functions, inference for misspecified models, and inference based on resampling methods. Robustification against microstructure noise is also of great theoretical and empirical interest (Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Xiu (2010)).

References


Note: This figure plots point estimates (solid) and 95% uniform confidence bands (dashed) of the one-step estimator for the linear regression model $Y_t = a + bV_t + \varepsilon_t$ and the ABC model $Y_t = a + bV_t^c + \varepsilon_t$ across all 25 subsample periods. The sampling interval is $\Delta = 1$ minute, and $k_n = [\Delta^{-2/5}]$.


Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.


